



Guia de EMR lançamento da Amazon

Amazon EMR



Amazon EMR: Guia de EMR lançamento da Amazon

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

As marcas comerciais e imagens comerciais da Amazon não podem ser usadas no contexto de nenhum produto ou serviço que não seja da Amazon, nem de qualquer maneira que possa gerar confusão entre os clientes ou que deprecie ou desprestige a Amazon. Todas as outras marcas comerciais que não são propriedade da Amazon pertencem aos respectivos proprietários, os quais podem ou não ser afiliados, estar conectados ou ser patrocinados pela Amazon.

Table of Contents

Sobre os EMR lançamentos da Amazon	1
Suporte padrão	2
Entendendo os EMR lançamentos da Amazon	2
Versionamento da versão	3
Política de suporte	4
Considerações	7
Versões de lançamento do Amazon EMR 7.x	9
Versões do aplicativo nas versões EMR 7.x da Amazon	9
emr-7.2.0	9
emr-7.1.0	65
emr-7.0.0	94
Versões de lançamento do Amazon EMR 6.x	132
Versões de aplicações nas versões 6.x do Amazon EMR	133
emr-6.15.0	134
emr-6.14.0	181
emr-6.13.0	229
emr-6.12.0	285
emr-6.11.1	350
emr-6.11.0	408
emr-6.10.1	447
emr-6.10.0	506
emr-6.9.1	548
emr-6.9.0	598
emr-6.8.1	646
emr-6.8.0	692
emr-6.7.0	739
emr-6.6.0	798
emr-6.5.0	859
emr-6.4.0	886
emr-6.3.1	919
emr-6.3.0	945
emr-6.2.1	978
emr-6.2.0	1005
emr-6.1.1	1038

emr-6.1.0	1059
emr-6.0.1	1085
emr-6.0.0	1102
Versões de lançamento do Amazon EMR 5.x	1124
Versões de aplicações nas versões 5.x do Amazon EMR	1127
emr-5.36.2	1128
emr-5.36.1	1155
emr-5.36.0	1201
emr-5.35.0	1238
emr-5.34.0	1264
emr-5.33.1	1289
emr-5.33.0	1318
emr-5.32.1	1340
emr-5.32.0	1364
emr-5.31.1	1392
emr-5.31.0	1411
emr-5.30.2	1434
emr-5.30.1	1453
emr-5.30.0	1475
emr-5.29.0	1498
emr-5.28.1	1517
emr-5.28.0	1535
emr-5.27.1	1555
emr-5.27.0	1572
emr-5.26.0	1591
emr-5.25.0	1611
emr-5.24.1	1630
emr-5.24.0	1647
emr-5.23.1	1666
emr-5.23.0	1682
emr-5.22.0	1701
emr-5.21.2	1720
emr-5.21.1	1736
emr-5.21.0	1754
emr-5.20.1	1773
emr-5.20.0	1789

emr-5.19.1	1809
emr-5.19.0	1825
emr-5.18.1	1843
emr-5.18.0	1860
emr-5.17.2	1877
emr-5.17.1	1893
emr-5.17.0	1909
emr-5.16.1	1927
emr-5.16.0	1943
emr-5.15.1	1960
emr-5.15.0	1976
emr-5.14.2	1992
emr-5.14.1	2008
emr-5.14.0	2024
emr-5.13.1	2042
emr-5.13.0	2057
emr-5.12.3	2073
emr-5.12.2	2088
emr-5.12.1	2104
emr-5.12.0	2119
emr-5.11.4	2136
emr-5.11.3	2151
emr-5.11.2	2166
emr-5.11.1	2181
emr-5.11.0	2197
emr-5.10.1	2213
emr-5.10.0	2228
emr-5.9.1	2244
emr-5.9.0	2259
emr-5.8.3	2276
emr-5.8.2	2291
emr-5.8.1	2305
emr-5.8.0	2320
emr-5.7.1	2336
emr-5.7.0	2351
emr-5.6.1	2366

emr-5.6.0	2381
emr-5.5.4	2396
emr-5.5.3	2411
emr-5.5.2	2425
emr-5.5.1	2440
emr-5.5.0	2454
emr-5.4.1	2470
emr-5.4.0	2485
emr-5.3.2	2500
emr-5.3.1	2514
emr-5.3.0	2529
emr-5.2.3	2544
emr-5.2.2	2558
emr-5.2.1	2573
emr-5.2.0	2588
emr-5.1.1	2603
emr-5.1.0	2617
emr-5.0.3	2632
emr-5.0.2	2646
emr-5.0.1	2659
emr-5.0.0	2673
Versões de lançamento do Amazon EMR 4.x	2688
Versões de aplicações nas versões 4.x do Amazon EMR	2689
Diferenças nas versões	2689
emr-4.9.6	2737
emr-4.9.5	2751
emr-4.9.4	2765
emr-4.9.3	2779
emr-4.9.2	2793
emr-4.9.1	2807
emr-4.8.5	2821
emr-4.8.4	2835
emr-4.8.3	2849
emr-4.8.2	2864
emr-4.8.1	2878
emr-4.8.0	2892

emr-4.7.4	2906
emr-4.7.3	2919
emr-4.7.2	2932
emr-4.7.1	2945
emr-4.7.0	2958
emr-4.6.1	2972
emr-4.6.0	2984
emr-4.5.0	2997
emr-4.4.0	3009
emr-4.3.0	3021
emr-4.2.0	3032
emr-4.1.0	3042
emr-4.0.0	3050
Versões 2.x e AMI 3.x	3057
Criar um cluster	3058
Instalar aplicações	3061
Personalizar configurações	3061
Hive	3068
HBase	3079
Pig	3092
Spark	3098
S3 DistCp	3101
O que há de novo?	3105
Amazon EMR 7.2.0	3105
Amazon EMR 6.15.0	3108
Amazon EMR 5.36.2	3127
Compatibilidade do SigV4	3131
Abordagem para CVE mitigar -2021-44228	3131
Solução de ação EMR bootstrap da Amazon para Log4j -2021-44228 e -2021-45046 CVE	
CVE	3132
Perguntas frequentes	3140
Arquivo	3142
Versão 6.14.0	3143
Versão 6.13.0	3163
Versão 6.12.0	3191
Versão 6.11.1	3228

Versão 6.11.0	3259
Versão 6.10.0	3271
Versão 6.9.0	3286
Versão 6.8.0	3307
Versão 6.7.0	3328
Versão 6.6.0	3361
Versão 5.35.0	3397
Versão 5.34.0	3401
Versão 6.5.0	3403
Versão 6.4.0	3406
Versão 5.32.0	3414
Versão 6.2.0	3419
Versão 5.31.0	3427
Versão 6.1.0	3433
Versão 6.0.0	3439
Versão 5.30.1	3444
Versão 5.30.0	3449
Versão 5.29.0	3454
Versão 5.28.1	3456
Versão 5.28.0	3457
Versão 5.27.0	3459
Versão 5.26.0	3462
Versão 5.25.0	3465
Versão 5.24.1	3468
Versão 5.24.0	3469
Versão 5.23.0	3471
Versão 5.22.0	3473
Versão 5.21.1	3476
Versão 5.21.0	3478
Versão 5.20.0	3480
Versão 5.19.0	3484
Versão 5.18.0	3485
Versão 5.17.1	3486
Versão 5.17.0	3487
Versão 5.16.0	3488
Versão 5.15.0	3489

Versão 5.14.1	3491
Versão 5.14.0	3491
Versão 5.13.0	3493
Versão 5.12.2	3494
Versão 5.12.1	3494
Versão 5.12.0	3494
Versão 5.11.3	3496
Versão 5.11.2	3496
Versão 5.11.1	3497
Versão 5.11.0	3497
Versão 5.10.0	3498
Versão 5.9.0	3500
Versão 5.8.2	3502
Versão 5.8.1	3503
Versão 5.8.0	3503
Versão 5.7.0	3505
Versão 5.6.0	3506
Versão 5.5.3	3506
Versão 5.5.2	3507
Versão 5.5.1	3507
Versão 5.5.0	3507
Versão 5.4.0	3509
Versão 5.3.1	3510
Versão 5.3.0	3510
Versão 5.2.2	3511
Versão 5.2.1	3511
Versão 5.2.0	3512
Versão 5.1.0	3513
Versão 5.0.3	3514
Versão 5.0.0	3514
Versão 4.9.5	3516
Versão 4.9.4	3516
Versão 4.9.3	3516
Versão 4.9.2	3517
Versão 4.9.1	3517
Versão 4.8.4	3518

Versão 4.8.3	3518
Versão 4.8.2	3519
Versão 4.8.0	3519
Versão 4.7.2	3520
Versão 4.7.1	3521
Versão 4.7.0	3521
Versão 4.6.0	3523
Versão 4.5.0	3525
Versão 4.4.0	3525
Versão 4.3.0	3527
Versão 4.2.0	3529
Configurar aplicações	3531
Configurar aplicações ao criar um cluster	3533
Fornecer uma configuração no console ao criar um cluster	3534
Forneça uma configuração usando o AWS CLI ao criar um cluster	3534
Forneça uma configuração usando o Java SDK ao criar um cluster	3535
Reconfigurar um grupo de instâncias de um cluster em execução	3535
Considerações ao reconfigurar um grupo de instâncias	3536
Reconfigurar um grupo de instâncias no console	3539
Reconfigure um grupo de instâncias usando o CLI	3540
Reconfigurar um grupo de instâncias usando o Java SDK	3545
Solução de problemas	3546
Armazenar dados de configuração sigilosos no AWS Secrets Manager	3549
Criar um segredo	3549
Conceda EMR à Amazon acesso para recuperar o segredo	3550
Usar o segredo em uma classificação de configuração	3550
Atualizar o valor do segredo	3551
Configurar aplicações para usar uma Máquina Virtual Java específica	3552
Considerações	3552
Substitua o JVM	3555
Portas de serviço	3557
Usuários do aplicativo	3559
Verificar dependências usando o repositório de artefatos	3560
EMRSistema de arquivos (EMRFS)	3563
Consistent view	3565
Habilitar a visualização consistente	3569

Entendendo como a visualização EMRFS consistente rastreia objetos no Amazon S3	3571
Lógica de novas tentativas	3572
EMRFSmetadados de visualização consistente	3573
Configure notificações de consistência para CloudWatch a Amazon SQS	3577
Configurar a visualização consistente	3578
EMRFSCLIREferência de comando	3583
Autorizando o acesso aos EMRFS dados no Amazon S3	3595
Criação de um provedor de credenciais personalizado para EMRFS dados no Amazon S3	3595
Gerenciando o AWS Security Token Service endpoint padrão	3597
Especificando a criptografia do Amazon S3 usando propriedades EMRFS	3598
Usando AWS KMS keys para EMRFS criptografia	3599
Criptografia do lado do servidor do Amazon S3	3600
Criptografia do lado do cliente do Amazon S3	3603
CloudWatch Agente da Amazon	3611
Criar um cluster	3613
Permissões obrigatórias	3613
Endpoint obrigatório	3614
Criar um cluster	3614
Métricas padrão	3615
Configuração	3618
Amazon EMR 7.1.0	3618
Amazon EMR 7.0.0	3651
Considerações	3657
Histórico	3657
Delta Lake	3659
Introdução	3659
Usar clusters do Delta Lake	3660
Delta Lake com Flink	3660
Delta Lake com Trino	3665
Delta Lake com Spark	3666
Delta Lake com Spark e Glue	3673
Considerações	3674
Histórico	3674
Flink	3676
Criar um cluster com o Flink	3679
Configurar o Flink	3680

Hive e Glue	3680
Arquivo de configuração	3682
Múltiplos nós primários	3683
Tamanho do processo de memória	3684
Tamanho do arquivo de saída de log	3685
Java 11	3686
Trabalhos do Flink	3691
Iniciar uma aplicação YARN do Flink como uma etapa em um cluster de execução prolongada	3691
Enviar o trabalho para uma aplicação Flink existente em um cluster de execução prolongada	3693
Enviar um trabalho transitório do Flink	3694
Shell do Flink Scala	3697
IU do Flink	3697
Escalador automático do Flink	3699
Visão geral	3699
Considerações	3699
Habilitação do escalador automático	3699
Configurações	3701
Otimização dos tempos de reinício	3706
Recuperação local de tarefas	3707
Pontos de verificação incrementais	3708
Recuperação refinada	3709
Reinício combinado	3710
Flink com Zeppelin	3711
Introdução	3711
Pré-requisitos	3711
Configurar o Zeppelin-Flink em um cluster do EMR	3711
Executar trabalhos do Flink com o Zeppelin-Flink em um cluster do EMR	3713
Histórico de versões do Flink	3718
Notas de lançamento do Flink por versão	3756
Ganglia	3757
Criar um cluster com o Ganglia	3759
Visualizar métricas do Ganglia	3760
Métricas do Hadoop e do Spark no Ganglia	3761
Histórico de versões do Ganglia	3762

Hadoop	3808
Configurar o Hadoop	3810
Configuração da tarefa	3810
Definições de configuração de daemon do Hadoop	4171
HDFSconfiguração	4441
Criptografia transparente HDFS na Amazon EMR	4442
Configurando a criptografia HDFS transparente	4443
Considerações sobre criptografia HDFS transparente	4446
Servidor de gerenciamento de chaves do Hadoop	4446
HDFSriptografia transparente em EMR clusters com vários nós principais	4450
Criar ou executar uma aplicação Hadoop	4452
Criar binários usando o Amazon EMR	4452
Processamento de dados com streaming	4455
Processe dados com um personalizado JAR	4460
Leia objetos restaurados	4463
Exemplos	4464
Considerações	4465
Ative o reconhecimento de acesso não uniforme à memória para YARN contêineres	4465
Histórico de versões do Hadoop	4468
Notas de versão do Hadoop por versão	4530
HBase	4536
Criar um cluster com o HBase	4540
Criação de um cluster HBase usando o console	4540
Criação de um cluster com HBase o uso do AWS CLI	4540
HBaseno Amazon S3 (modo de armazenamento Amazon S3)	4541
Habilitando HBase no Amazon S3	4543
Usar um cluster de réplica de leitura	4543
HFileMonitoramento persistente	4545
Considerações operacionais	4547
Registros de gravação antecipada (WAL) para Amazon EMR	4551
WALespaços de trabalho	4553
Permissões obrigatórias	4553
Habilitando WAL	4554
Restaurando de WAL	4557
Configurações de segurança	4558
Usando AWS PrivateLink	4558

WALpreços e métricas	4560
Marcação de espaços de trabalho WAL	4561
Considerações e disponibilidade	4563
EMRWALCLIreferência	4564
Usando a HBase concha	4567
Criar uma tabela	4568
Inserir um valor	4568
Obter um valor	4568
Excluir uma tabela	4568
Acesse HBase tabelas com o Hive	4569
Usando HBase instantâneos	4570
Criar um snapshot usando uma tabela	4571
Excluir um snapshot	4571
Exibir informações do snapshot	4571
Exportar um snapshot para o Amazon S3	4572
Importar um snapshot do Amazon S3	4572
Restaurar uma tabela a partir de instantâneos dentro do shell HBase	4573
Configurar HBase	4574
Alterações na alocação de memória em YARN	4575
HBasenúmeros de porta	4576
HBaseconfigurações do site para otimizar	4577
Visualize a interface HBase do usuário	4579
Exibir arquivos de log do HBase	4581
Monitor HBase com Ganglia	4581
Migrando de versões anteriores HBase	4583
HBasehistórico de lançamentos	4583
HCatalog	4656
Criar um cluster com o HCatalog	4657
Usando HCatalog	4658
Desative a gravação direta ao usar HCatalog HStorer	4658
Crie uma tabela usando o HCat CLI e use esses dados no Pig	4659
Acessando a tabela usando o Spark SQL	4660
Exemplo: Crie uma HCatalog tabela e grave nela usando o Pig	4662
HCataloghistórico de lançamentos	4663
Hive	4732
Diferenças e considerações sobre o Hive na Amazon EMR	4735

Diferenças entre o Apache Hive na Amazon EMR e o Apache Hive	4735
Diferenças no Hive entre as EMR versões 4.x e 5.x da Amazon	4736
Recursos adicionais do Hive na Amazon EMR	4737
Configurar um metastore externo para o Hive	4742
Usando o AWS Glue Data Catalog como metastore para o Hive	4743
Usando um SQL banco de dados externo My ou Amazon Aurora	4750
Usar o driver JDBC do Hive	4753
Melhorar a performance do Hive	4755
Habilitando o committer otimizado do Hive EMRFS S3	4756
Usando o S3 Select	4757
MSCKOtimização	4759
Use o Hive LLAP	4760
Habilitar o LLAP	4761
Comece LLAP no seu cluster	4762
Verifique o LLAP status	4763
Iniciar ou parar LLAP	4763
Redimensionar contagem de LLAP daemons	4763
Criptografia no Hive	4763
Criptografia modular em Parquet no Hive	4763
Criptografia em trânsito em HS2	4767
Histórico de versões do Hive	4769
Notas da versão do Hive por versão	4839
Hudi	4921
Como o Hudi funciona	4923
Noções básicas sobre os tipos de armazenamento de conjuntos de dados: copiar na gravação ou mesclar na leitura	4923
Registrar um conjunto de dados do Hudi com seu metastore	4925
Considerações e limitações	4925
Criar um cluster com o Hudi instalado	4927
Trabalhe com um conjunto de dados do Hudi	4928
Inicializar uma sessão do Spark para Hudi	4931
Gravar em um conjunto de dados do Hudi	4932
Upsert dados	4937
Excluir um registro	4938
Ler em um conjunto de dados do Hudi	4940
Use o Hudi CLI	4942

Histórico de versões do Hudi	4942
Hue	4946
Versões do Hue	4946
Recursos compatíveis e não suportados do Hue na Amazon EMR	4948
Considerações	4949
Desempenho com grandes tabelas de metadados do Hue	4949
Incompatibilidade entre as versões do Hue	4951
Conectar-se à interface de usuário da Web do Hue	4951
Usando o Hue com um banco de dados remoto na Amazon RDS	4952
Solução de problemas	4954
Configurações avançadas do Hue	4955
Configurar o Hue para usuários LDAP	4955
Histórico de versões do Hue	4958
Iceberg	5021
Como o Iceberg funciona	5022
Usar um cluster com o Iceberg	5024
Usar um cluster do Iceberg com o Spark	5024
Usar um cluster do Iceberg com o Trino	5029
Usar um cluster do Iceberg com o Flink	5031
Usar um cluster do Iceberg com o Hive	5036
Considerações e limitações	5039
Considerações sobre o uso do Iceberg com o Spark	5039
Considerações sobre o uso do Iceberg com o Trino	5039
Considerações sobre o uso do Iceberg com o Flink	5040
Considerações sobre o uso do Iceberg com o Hive	5040
Histórico de versões do Iceberg	5040
Notas da versão do Iceberg por versão	5042
Bloco de anotações Jupyter	5044
EMR Studio	5044
Bloco de anotações do EMR	5044
JupyterHub	5045
Crie um cluster com JupyterHub	5049
Considerações ao usar JupyterHub na Amazon EMR	5051
Configurando JupyterHub	5052
Configurar a persistência de cadernos no Amazon S3	5053
Conectar-se ao nó principal e aos servidores de cadernos	5054

JupyterHub configuração e administração	5055
Adicionar usuários e administradores do Caderno Jupyter	5057
Instalar Kernels e bibliotecas adicionais	5068
JupyterHub histórico de lançamentos	5073
Livy	5109
Habilitando HTTPS	5111
Histórico de versões do Livy	5112
MXNet	5156
MXNethistórico de lançamentos	5158
Oozie	5187
Usando o Oozie com um banco de dados remoto na Amazon RDS	5189
Configurar a versão do Java para Oozie	5192
Histórico de versões do Oozie	5193
Notas de lançamento do Oozie por versão	5250
Phoenix	5251
Criar um cluster com o Phoenix	5254
Personalizar as configurações do Phoenix	5255
Clientes do Phoenix	5256
Histórico de versões do Phoenix	5259
Pig	5334
Enviar um trabalho do Pig	5336
Envie trabalhos do Pig usando o console da Amazon EMR	5337
Envie o trabalho do Pig usando o AWS CLI	5338
Chamar funções definidas pelo usuário no Pig	5339
Chame JAR arquivos do Pig	5340
Chamar scripts em Python/Jython no Pig	5340
Histórico de versões do Pig	5341
Presto e Trino	5407
Usando o Presto com o AWS Glue Data Catalog	5410
Especificando o AWS Glue Data Catalog como metastore	5411
IAMpermissões	4746
Considerações ao usar o Catálogo de Dados do AWS Glue	5415
Usando o S3 Select Pushdown	5416
O S3 Select Pushdown é adequado para minha aplicação?	5416
Considerações e limitações	5417
Habilitar o S3 Select Pushdown com o PrestoDB ou Trino	5417

Adicionar conectores de banco de dados	5418
UsandoSSL/TLS e LDAPS	5419
Usando a LDAP autenticação	5421
Ativar o modo estrito do Presto	5428
Considerações	5430
Lidar com a perda de instâncias spot no Presto	5431
Execução tolerante a falhas	5433
Configuração	5433
Gerenciador de trocas	5433
Considerações e limitações	5435
Usar a escalabilidade automática do Presto com desativação tranquila	5435
Considerações sobre o Presto na Amazon EMR	5437
Executável da linha de comando do Presto	5437
Propriedades de implantação do Presto não configuráveis	5437
Instalação do PrestoDB e do Trino	5438
EMRFS e configuração do PrestoOS3 FileSystem	5439
Configuração padrão para personificação do usuário final	5440
Porta padrão para interface Web do Presto	5440
Problema com a execução do bucket do Hive em algumas versões	5440
Histórico de versões do Presto	5441
Notas de lançamento do Trino (PrestoSQL) por versão	5506
Spark	5509
Criação de um cluster do Spark	5513
Execute aplicativos Spark com o Docker na Amazon 6.x EMR	5515
Considerações	5516
Criar uma imagem do Docker	5517
Usando imagens Docker da Amazon ECR	5518
Use o AWS Glue Data Catalog como metastore para o Spark SQL	5523
Especificando o AWS Glue Data Catalog como metastore	5524
IAMpermissões	4746
Considerações	4748
Configurar o Spark	5529
Padrões do Spark	5530
Configuração da coleta de resíduos	5531
maximizeResourceAllocation	5532
Comportamento de desativação	5534

Variável de ThriftServer ambiente Spark	5538
Alterar as configurações padrão do Spark	5538
Migrar do Apache Log4j 1.x para Log4j 2.x	5541
Otimizar a performance do Spark	5541
Execução adaptável de consultas	5542
Corte de partição dinâmico	5544
Nivelamento de subconsultas escalares	5546
DISTINCTantes INTERSECT	5547
Junção com filtro Bloom	5548
Reordenação de junção otimizada	5549
Armazenamento em Cache de Fragmentos Resultantes	5549
Habilitar o Armazenamento em Cache de Fragmentos Resultantes do Spark	5550
Considerações	5551
Use o RAPIDS acelerador	5553
Selecionar tipos de instância	5553
Definição das configurações da aplicação	5553
Adição de uma ação de bootstrap	5561
Iniciar o cluster	5562
Acessar o shell do Spark	5562
Use o Amazon SageMaker Spark para aprendizado de máquina	5564
Escrever uma aplicação do Spark	5564
Scala	5564
Java	5565
Python	5566
Melhora da performance do Spark com o S3	5568
User S3 Select	5568
EMRFSCommitter otimizado para S3	5573
Use o protocolo de confirmação EMRFS otimizado para S3	5580
Tentar novamente solicitações do S3	5586
Adicionar uma etapa do Spark	5589
Substituir as definições de configuração padrão do Spark	5593
Visualizar o histórico de aplicações do Spark	5593
Acesse a web do Spark UIs	5594
Usando o Spark com o Amazon Kinesis Data Streams	5594
Usar o Spark no Amazon Redshift	5595
Iniciar uma aplicação do Spark	5595

Autenticação no Amazon Redshift	5596
Leitura e gravação para o Amazon Redshift	5598
Considerações	5600
Histórico de versões do Spark	5601
Sqoop	5665
Versões do Sqoop	5665
Considerações sobre o Sqoop na Amazon EMR	5667
Usando o Sqoop com integração HCatalog	5667
Suporte para Sqoop JDBC e banco de dados	5667
Proteção da senha	5669
Histórico de versões do Sqoop	5670
TensorFlow	5724
TensorFlow compilações por tipo de EC2 instância da Amazon	5726
Segurança	5726
Usando TensorBoard	5726
TensorFlow histórico de lançamentos	5727
Tez	5750
Criar um cluster com o Tez	5752
Configurar o Tez	5753
Exemplo de configuração	5753
Abertura de divisão assíncrona do Tez	5754
Interface do usuário da Web do Tez	5756
Timeline Server	5757
Histórico de versões do Tez	5757
Notas da versão do Tez por versão	5801
Zeppelin	5808
Considerações ao usar o Zeppelin na Amazon EMR	5810
Histórico de versões do Zeppelin	5811
ZooKeeper	5869
ZooKeeper histórico de lançamentos	5871
Conectores e utilitários	5910
Exportar, consultar e unir tabelas no DynamoDB	5910
Configurar uma tabela do Hive para executar comandos do Hive	5912
Exemplos de comandos do Hive para exportar, importar e consultar dados	5921
Otimizar o desempenho	5930
Kinesis	5934

O que posso fazer com a integração entre a Amazon EMR e o Amazon Kinesis?	5934
Análise com ponto de verificação de fluxos do Amazon Kinesis	5935
Considerações sobre a performance	5937
Agende análises do Amazon Kinesis com a Amazon EMR	5937
Migração do conector Spark Kinesis para SDK 2.x para Amazon 7.0 EMR	5937
S3 DistCp (s3-dist-cp)	5946
Opções do S3 DistCp	5947
Adicionando o S3 DistCp como uma etapa em um cluster	5955
Limpando após falhas em trabalhos do S3 DistCp	5957
Executar comandos e scripts em um cluster	5959
Enviar uma etapa JAR personalizada para executar um script ou comando	5960
Outras formas de usar <code>command-runner.jar</code>	5961
Glossário do AWS	5963
.....	5964

Sobre os EMR lançamentos da Amazon

Uma EMR versão da Amazon é um conjunto de aplicativos de código aberto do ecossistema de big data. Cada versão inclui diferentes aplicativos, componentes e recursos de big data que você seleciona para que a Amazon EMR instale e configure ao criar um cluster. Os aplicativos são empacotados usando um sistema baseado no [Apache BigTop](#), que é um projeto de código aberto associado ao ecossistema Hadoop. Este guia fornece informações sobre os aplicativos incluídos nas EMR versões da Amazon.

Para obter mais informações sobre como começar e trabalhar com a AmazonEMR, consulte o [Guia EMR de gerenciamento da Amazon](#).

Ao iniciar um cluster, você pode escolher entre várias versões da AmazonEMR. Isso permite que você teste e use versões de aplicações que atendem aos seus requisitos de compatibilidade. Você especifica o número da versão com o rótulo da versão. Os rótulos de versão estão no formato `emr-x.x.x`. Por exemplo, `emr-7.2.0`.

A partir do Amazon EMR 5.18.0, você pode usar o repositório de EMR artefatos da Amazon para criar seu código de trabalho com base nas versões exatas das bibliotecas e dependências que estão disponíveis em versões específicas da Amazon. Para ter mais informações, consulte [Verificar dependências usando o repositório de artefatos do Amazon EMR](#).

Para receber atualizações quando um novo EMR lançamento da Amazon estiver disponível, assine o [RSSfeed para ver as notas de EMR lançamento da Amazon](#).

Detalhes da versão mais recente, incluindo versões de aplicativos, notas de lançamento, componentes e classificações de configuração das séries EMR 7.x, 6.x e 5.x da Amazon:

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

Note

Novos EMR lançamentos da Amazon são disponibilizados em diferentes regiões por um período de vários dias, começando com a primeira região na data de lançamento inicial. A versão mais recente pode não estar disponível em sua região durante esse período.

Notas de lançamento dos últimos EMR lançamentos da Amazon e um histórico de todos os lançamentos:

- [O que há de novo?](#)
- [EMR Arquivo de notas de lançamento da Amazon](#)

Um histórico abrangente das versões do aplicativo em cada EMR versão da Amazon:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Detalhes de cada EMR versão da Amazon e diferenças entre as séries de lançamento, quando aplicável:

- [Versões de lançamento do Amazon EMR 7.x](#)
- [Versões de lançamento do Amazon EMR 6.x](#)
- [Versões de lançamento do Amazon EMR 5.x](#)
- [Versões de lançamento do Amazon EMR 4.x](#)
- [Versões Amazon EMR 2.x e AMI 3.x](#)

Suporte EMR padrão da Amazon

Entendendo os EMR lançamentos da Amazon

A EMR versão da Amazon é o veículo que fornece o software necessário para executar seus aplicativos de código aberto em EC2EKS, ou plataformas sem servidor. Uma EMR versão da Amazon é composta por três partes: Runtime Environment, Core Engines e Extras.

- Ambiente de tempo de execução — O ambiente de tempo de execução inclui o sistema operacional incluído como parte da Amazon Machine Image (AMI) ou outra imagem de contêiner quando você inicia seu cluster ou contêiner (por exemplo, Amazon Linux 2023). Também inclui os tempos de execução da linguagem, como o Amazon JDK Corretto, junto com outras ferramentas relacionadas que a imagem instala.

- Mecanismos principais — incluem o software principal de código aberto, como o Apache Hive, e os formatos de tabela aberta, como o Apache Iceberg.
- Extras — Isso inclui bibliotecas de conveniência e pacotes Python, como software de código aberto, como `mariaadb-connector-java` o Apache Pig.

A versão de EMR lançamento da Amazon que você seleciona agrupa todos esses componentes em uma nova versão do Amazon Machine Images () baseado no Amazon Linux (ALAMI) ou imagens de contêiner. A Amazon EMR mantém a mesma versão AL para todos os nós de um cluster, EKS contêiner ou aplicativo sem servidor. A Amazon EMR pretende obter o ambiente de execução mais recente dentro de 90 dias após seu lançamento pelo Amazon Linux e novas versões de código aberto do Core Engines dentro de 90 dias a partir do lançamento upstream. Os extras são liberados conforme necessário. Você pode encontrar a lista do software nas notas de lançamento de cada EMR versão da Amazon.

Versionamento da versão

Para ajudar você a entender o escopo das mudanças em cada versão, a Amazon EMR usa versionamento semântico. O controle de versão semântico segue um formato de `<major>.<minor>.<patch>` Como exemplo, vamos considerar a versão 7.0.0.

O primeiro dígito indica a versão principal, ou a versão principal. Os principais lançamentos geralmente trazem mudanças, melhorias e novos recursos substanciais para o Runtime Environment ou Core Engine que não são compatíveis com versões anteriores. Por exemplo, a versão principal 7.x usa o Amazon Linux 2023 com o Amazon Corretto JDK 17 como padrão. Como resultado, a versão contém várias alterações importantes em comparação com a versão principal 6.x, que vem no Amazon Linux 2 e com o Amazon Corretto JDK 8 como padrão.

O segundo dígito indica o lançamento da versão secundária, ou versão secundária. Versões secundárias são versões não compatíveis com versões anteriores que contêm alterações incrementais, melhorias e recursos nos Core Engines e Extras, além de novas funcionalidades. Por exemplo, o 6.15 usa o Apache Spark 3.4.1, em comparação com o 6.11.1, que vem com o Apache Spark 3.3.2.

O terceiro dígito indica o lançamento de uma versão de patch, ou lançamento de patch. As versões de patch devem ser versões compatíveis com versões anteriores que contêm correções e atualizações de patches para o Core Engine e os extras, mas sem novas funcionalidades ou atualizações de OSS versão. Por exemplo, 6.11.1 e 6.11.0 contêm a mesma versão 3.3.2 do Apache

Spark. OSS A partir da versão 6.6.x, a versão mais recente do patch oferece o patch mais recente do Runtime Environment disponível. O Runtime Environment não afeta o lançamento de patches da AmazonEMR. Por exemplo, 6.11.1 foi lançado com a AL versão 2.0.20240109.0 de 19 de dezembro de 2023 a 10 de janeiro de 2024, mas com a AL versão 2.0.20240124.0 a partir de 11 de janeiro de 2024. Alguns EKS lançamentos da Amazon EMR on permitem que você adicione tags ou sufixos às versões semânticas para fornecer funcionalidades alternativas. Por exemplo, o `emr-6.15.0-latest` lançamento do Amazon EMR on EKS lança uma versão de lançamento JDK habilitada para o Amazon Corretto 8, enquanto o `emr-6.15.0-java17-latest` lançamento do Amazon EMR on EKS permite que você lance uma versão de lançamento habilitada para o Amazon Corretto 17JDK.

Política de suporte

Introdução

EMRA Amazon tem como objetivo emitir novos lançamentos menores no mínimo uma vez a cada 90 dias e fornecer suporte para lançamentos menores por um período de 24 meses a partir da data de lançamento. Esse suporte abrange o Runtime Environment e os Core Engines com suas dependências associadas, e não abrange Extras ou suas dependências associadas. O Runtime Environment também inclui componentes proprietários da plataforma e APIs que são necessários para iniciar, parar e operar os clusters e os aplicativos. Esse cronograma de lançamento consistente garante um ciclo previsível, tornando mais conveniente planejar, testar e fazer a transição para uma versão que forneça suporte e segurança.

O que esperar do Standard Support

O Standard Support fornece correções em tíquetes de suporte técnico para problemas que você encontra nos componentes do Runtime Environment e do Core Engines nas configurações recomendadas. Todas as correções estão sujeitas à disponibilidade. [A configuração recomendada é o uso da Amazon EMR sem nenhuma modificação, acréscimo ou alteração nos binários e configurações presentes em uma EMR versão da Amazon, exceto aqueles encontrados na documentação da Amazon. EMR](#) A Amazon EMR implanta correções no patch mais recente, na versão secundária ou principal da versão da Amazon EMR dentro de 90 dias após a verificação da correção. A Amazon aplica correções EMR automaticamente quando você lança um novo EMR no EC2 cluster, lança um novo Amazon EMR no EKS contêiner ou aciona um novo trabalho EMR sem servidor. Componentes extras são fornecidos por conveniência e a Amazon EMR não fornece correções relacionadas a Extras.

Componentes

O Standard Support abrange os componentes Runtime Environment e Core Engines — por exemplo, sistema operacional, tempos de execução de linguagem e software de código aberto básico, como Apache Hive e Apache Iceberg. Você pode encontrar a lista completa dos componentes compatíveis com cada versão no [Sobre os EMR lançamentos da Amazon Amazon Linux FAQs](#).

A lista a seguir descreve o suporte que oferecemos para diferentes tipos de componentes em Standard Support:

- **Componentes do Runtime Environment:** os componentes do Runtime Environment receberão correções nos tíquetes de suporte técnico. As correções são categorizadas como (a) bugs críticos, (b) problemas críticos de corrupção de dados e (c) problemas críticos de segurança. Quando qualificada, a Amazon EMR transferirá as correções do Runtime Environment para versões mais antigas. Para compatibilidade operacional com os componentes de código aberto do Core Engine, certos componentes do Runtime Environment devem permanecer em versões específicas para evitar a interrupção dos aplicativos do cliente. Para esses componentes, a Amazon EMR depende do código aberto upstream para a disponibilidade de correções. Quando as correções estiverem disponíveis em código aberto, forneceremos a versão estável mais recente dentro de 90 dias após a verificação pela AmazonEMR.
- **Componentes do Core Engines:** O Core Engines fornece as versões mais recentes para muitos projetos de código aberto, cada um com centenas de bibliotecas de dependências transitivas. Embora as comunidades de código aberto que gerenciam esses projetos tentem resolver problemas e vulnerabilidades e exposições comuns conhecidas (CVE) com frequência, as versões mais recentes ainda podem conter bugs conhecidos e CVEs EMRA Amazon depende do código aberto upstream para disponibilidade de correções e fornecerá a versão estável mais recente como parte dos componentes do Core Engine em até 90 dias após a verificação pela Amazon. EMR Em certos casos, a Amazon EMR pode fornecer uma correção para um CVE dos componentes do Core Engine que precisa ser tratada antes do código aberto upstream. A Amazon EMR também fornece suporte técnico e correções em recursos que são adicionados ao código aberto nas configurações recomendadas. Não transferimos correções de componentes do Core Engine para versões mais antigas de Patch ou Minor.
- **Componentes extras:** EMR a Amazon não oferece suporte a componentes extras. Componentes extras são projetos de código aberto fornecidos por conveniência e EMR a Amazon não fornece correções para problemas encontrados com eles. Quaisquer solicitações ou correções de suporte podem ser atendidas por meio da comunidade de código aberto que oferece suporte a esses componentes.

Ciclo de vida do Standard Support

A seguir, descrevemos os marcos do ciclo de vida do Standard Support:

- Suporte padrão: os EMR lançamentos da Amazon são elegíveis para o Suporte Padrão 24 meses a partir da data de lançamento. Você pode criar tíquetes de suporte técnico e esperar atualizações para os problemas encontrados nesses lançamentos.
- Fim do suporte: após o término do suporte padrão, os EMR lançamentos da Amazon entram no estágio End of Support (eOS) por 12 meses. As versões do EoS não estão qualificadas para suporte técnico e você não poderá criar nenhum ticket para clusters, contêineres ou trabalhos em execução nessas versões. As versões do EoS não receberão correções, patches ou atualizações. As versões do EoS serão removidas do console, mas continuarão disponíveis por meio do API e AWS CLI. Você sempre pode continuar executando cargas de trabalho nas versões do EoS. É altamente recomendável que você migre para a EMR versão mais recente da Amazon para que continue recebendo patches de segurança, permaneça qualificado para suporte técnico e possa criar tickets de suporte quando necessário.
- Fim da vida útil: após o prazo de End of Support (EoS), as versões são consideradas End of Life (EoL). Embora você possa continuar executando clusters de EoL, a Amazon EMR se reserva o direito de remover as versões de EoL do API e case-by-case com base SDK em questões operacionais e de segurança. É altamente recomendável que você migre para a versão mais recente da AmazonEMR, pois as versões EoL podem ser removidas da API e SDK em casos excepcionais.

Suporte do Bridge

A Amazon EMR anunciou essa nova política de suporte em 25 de julho de 2024. De acordo com essa política, as versões da Amazon EMR lançadas em ou antes de 24 de julho de 2022 agora são designadas como End of Support. No entanto, para que você tenha mais tempo para planejar e migrar para versões mais novas, a Amazon EMR oferecerá o Bridge Support, equivalente ao Standard Support, para essas versões mais antigas lançadas no período de dois anos antes deste anúncio.

Depois de 25 de julho de 2024, você pode consultar o status atual e os cronogramas de suporte nas notas de versão.

A tabela a seguir mostra o status do suporte para todos os EMR lançamentos existentes da Amazon no momento do anúncio da política, 25 de julho de 2024:

Lançamentos e períodos de suporte

Versão de EMR lançamento da Amazon	Data da versão inicial	Data de término do suporte padrão	Data de início do fim do suporte	Data de início do fim da vida útil
7.2.0	25 de julho de 2024	24 de julho de 2026	25 de julho de 2026	25 de julho de 2027
7.1.0	23 de abril de 2024	22 de abril de 2026	23 de abril de 2026	23 de abril de 2027
7.0.0	19 de dezembro de 2023	18 de dezembro de 2025	19 de dezembro de 2025	19 de dezembro de 2026
5.36.x e 6.6.x — 6.15.x	9 de maio de 2022 a 13 de novembro de 2023	Suporte do Bridge até 24 de janeiro de 2026	25 de janeiro de 2026	25 de janeiro de 2026
<ul style="list-style-type: none"> Série 6.x: 6.5.0 e inferior Série 5.x: 5.35.0 e inferior Séries 4.x, 3.x e 2.x 	1º de janeiro de 2013 a 30 de março de 2022	Suporte do Bridge até 24 de julho de 2025	25 de julho de 2025	25 de julho de 2025

Considerações

O Standard Support está disponível para todos os modelos de EMR implantação da Amazon (EMREMRon EC2EKS, Amazon on e EMR Serverless), em todas as regiões onde a Amazon EMR está disponível, sem custo adicional. Os clusters executados com as configurações recomendadas são automaticamente qualificados para suporte, conforme descrito na política, portanto, você não precisa realizar nenhuma ação adicional para ativar o suporte.

- O Standard Support oferece suporte apenas aos componentes necessários para EMR clusters da Amazon. A Amazon não EMR pode garantir a disponibilidade de patches e correções

de segurança no caso de os componentes de código aberto do Core Engine atingirem o EoL upstream ou quando as atualizações de segurança não estiverem mais disponíveis para dependências. Embora você possa optar por instalar os Extras, a Amazon não EMR oferecerá suporte a eles ou a suas dependências. Por exemplo, você pode instalar aplicativos de terceiros em sua personalização AMI para reforçar a segurança do seu cluster, instalar componentes adicionais ou copiar objetos usando scripts de ação de bootstrap ou SSH em seu cluster e atualizar as versões padrão do pacote. A Amazon EMR não oferece suporte a esses componentes. O Standard Support não cobre ações de bootstrap, pacotes, bibliotecas, seu código personalizado e aplicativos bring-your-own personalizados fornecidos pelo cliente que você pode configurar EMR a Amazon para instalar para sua conveniência.

- Seus clusters existentes não serão afetados, independentemente da EMR versão da Amazon que estejam executando. Você pode continuar executando os clusters existentes sem interrupções. Você também pode continuar lançando novos clusters e executando trabalhos em qualquer uma das versões existentes e novas. Todos os lançamentos existentes e novos lançamentos no momento da entrada em vigor da política são cobertos pelo Standard Support por 24 meses a partir da data inicial de lançamento da versão de EMR lançamento da Amazon. A Amazon EMR fornecerá suporte rápido durante o anúncio inicial da política. Para receber suporte ininterrupto, recomendamos testar seus aplicativos e atualizá-los imediatamente para a versão mais recente da Amazon. EMR
- A Amazon EMR não alterará os componentes do Standard Support em versões ou clusters existentes. No entanto, a Amazon EMR se reserva o direito de honrar o fim da vida útil inicial e remover esses componentes nas novas versões ou transferir a versão existente para o status End of Support (EoS) ou End of Life (EoL) em casos excepcionais. case-by-case Notificaremos você sobre qualquer remoção por meio dos canais disponíveis.
- Quando uma nova região for lançada, a Amazon só EMR oferecerá suporte às versões de EMR lançamento da Amazon sob o Suporte Padrão, lançadas seis meses antes da data em que a nova região se tornará disponível ao público em geral.
- A Amazon EMR não atualizará automaticamente seus clusters existentes para as versões mais recentes. No entanto, você pode optar por atualizar os novos clusters para as versões mais recentes do patch, se desejar.
- O custo do EMR uso não suportado da Amazon será contabilizado na sua AWS fatura. Mesmo se você estiver usando a Amazon de uma EMR forma não suportada, os custos associados a esse uso ainda fazem parte do seu AWS consumo geral e serão incluídos no cálculo de suas taxas de suporte.

Para obter mais informações, entre em contato com o [AWS Developer Support](#).

Versões de lançamento do Amazon EMR 7.x

Esta seção contém versões de aplicativos, notas de lançamento, versões de componentes e classificações de configuração disponíveis em cada versão de lançamento do Amazon EMR 7.x.

Ao iniciar um cluster, você pode escolher entre várias versões da AmazonEMR. Isso permite que você teste e use versões de aplicações que atendem aos seus requisitos de compatibilidade. Você especifica o número da versão com o rótulo da versão. Os rótulos de versão estão no formato `emr-x.x.x`. Por exemplo, `emr-7.2.0`.

Novos EMR lançamentos da Amazon são disponibilizados em diferentes regiões por um período de vários dias, começando com a primeira região na data de lançamento inicial. A versão mais recente pode não estar disponível em sua região durante esse período.

Para obter uma tabela abrangente das versões do aplicativo em cada versão do Amazon EMR 7.x, consulte [Versões do aplicativo nas versões EMR 7.x da Amazon](#).

Tópicos

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Amazon EMR versão 7.2.0](#)
- [Amazon EMR versão 7.1.0](#)
- [Amazon EMR versão 7.0.0](#)

Versões do aplicativo nas versões EMR 7.x da Amazon

Para obter uma tabela abrangente que lista as versões do aplicativo disponíveis em cada versão do Amazon EMR 7.x, abra [Versões do aplicativo nas versões do Amazon EMR 7.x](#) em seu navegador.

Amazon EMR versão 7.2.0

Ciclo de vida compatível com 7.2.0

A tabela a seguir descreve as datas de ciclo de vida suportadas para a EMR versão 7.2.0 da Amazon.

Fase de Support	Data
Data da versão inicial	25 de julho de 2024
Suporte padrão até	24 de julho de 2026
Término do suporte	25 de julho de 2026
Fim da vida útil	25 de julho de 2027

Para obter uma lista de vulnerabilidades e exposições (CVEs) comuns críticas e altas que não afetam os EMR clusters sob o software e as configurações recomendados, consulte [7.2.0 vulnerabilidades e exposições comuns conhecidas](#) para mecanismos principais. Para ver uma lista dos CVEs que foram corrigidos na versão 7.2.0, consulte [7.2.0 corrigiu vulnerabilidades e exposições comuns](#).

7.2.0 vulnerabilidades e exposições comuns

Esta página lista todas as vulnerabilidades e exposições comuns (CVEs) relacionadas ao Amazon 7.2.0. EMR

7.2.0 vulnerabilidades e exposições comuns conhecidas para mecanismos principais

A tabela a seguir lista todos os CVEs que não afetam os EMR clusters executados na Amazon EMR 7.2.0. EMRA Amazon depende do código aberto upstream para disponibilidade de correções e fornecerá a versão estável mais recente como parte dos componentes do Core Engine em até 90 dias após a EMR verificação das correções pela Amazon.

CVEID	Gravidade	CVEdetalhes URL
CVE-2024-1597	Crítico	https://nvd.nist.gov/vuln/detail/CVE-2024-1597
CVE-2023-44981	Crítico	https://nvd.nist.gov/vuln/detail/CVE-2023-44981
CVE-2023-310 47	Crítico	https://nvd.nist.gov/vuln/detail/CVE-2023-31047

CVEID	Gravidade	CVEdetalhes URL
CVE-2023-2568	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2023-25668
CVE-2023-2564	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2023-25664
CVE-2022-46337	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2022-46337
CVE-2022-4289	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2022-42889
CVE-2022-39135	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2022-39135
CVE-2022-37865	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2022-37865
CVE-2022-26612	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2022-26612
CVE-2022-25 168	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2022-25168
CVE-2022-1471	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2022-1471
CVE-2021-37404	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2021-37404
CVE-2021-35958	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2021-35958
CVE-2021-327 98	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2021-32798
CVE-2021-32797	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2021-32797

CVEID	Gravidade	CVEdetalhes URL
CVE-2020-9548	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2020-9548
CVE-2020-9547	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2020-9547
CVE-2020-9546	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2020-9546
CVE-2020-8840	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2020-8840
CVE-2019-20 445	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2019-20445
CVE-2019-20 44	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2019-20444
CVE-2019-20330	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2019-20330
CVE-2019-17 531	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2019-17531
CVE-2019-17 267	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2019-17267
CVE-2019-17 195	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2019-17195
CVE-2019-16 943	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2019-16943
CVE-2019-16 942	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2019-16942
CVE-2019-16 335	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2019-16335

CVEID	Gravidade	CVEdetalhes URL
CVE-2019-14 893	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2019-14893
CVE-2019-14 892	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2019-14892
CVE-2019-14 540	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2019-14540
CVE-2019-14 379	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2019-14379
CVE-2019-10 202	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2019-10202
CVE-2018-7 489	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2018-7489
CVE-2018-1936 2	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2018-19362
CVE-2018-1936 1	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2018-19361
CVE-2018-19360	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2018-19360
CVE-2018-14 721	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2018-14721
CVE-2018-14 720	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2018-14720
CVE-2018-14 719	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2018-14719
CVE-2018-14 718	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2018-14718

CVEID	Gravidade	CVEdetalhes URL
CVE-2018-11 307	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2018-11307
CVE-2017-7658	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2017-7658
CVE-2017-7 657	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2017-7657
CVE-2017-7525	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2017-7525
CVE-2017-17485	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2017-17485
CVE-2017-15095	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2017-15095
CVE-2017-12 629	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2017-12629
CVE-2016-5018	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2016-5018
CVE-2015-1832	Crítico	https://nvd.nist.gov/vuln/detail/CVE -2015-1832
CVE-2024-2913	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2024-29133
CVE-2024-29 131	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

CVEID	Gravidade	CVEdetalhes URL
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
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-2023-3447	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2023-34478

CVEID	Gravidade	CVEdetalhes URL
CVE-2023-3445	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2023-34455
CVE-2023-3445	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2023-34454
CVE-2023-3445 3	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
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

CVEID	Gravidade	CVEdetalhes URL
CVE-2023-25670	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2023-25670
CVE-2023-2569	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2023-25669
CVE-2023-2567	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2023-25667
CVE-2023-2566	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2023-25666
CVE-2023-25665	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2023-25665
CVE-2023-2563	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2023-25663
CVE-2023-2562	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
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

CVEID	Gravidade	CVEdetalhes URL
CVE-2023-23969	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2023-23969
CVE-2023-2083	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-2022-46751	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-46751
CVE-2022-45693	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-45693
CVE-2022-45685	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-45685
CVE-2022-44729	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-44729
CVE-2022-42969	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-42969
CVE-2022-42890	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-42890
CVE-2022-42004	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-42004
CVE-2022-42003	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-42003
CVE-2022-41704	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2022-41704

CVEID	Gravidade	CVEdetalhes URL
CVE-2022-4089	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2022-40899
CVE-2022-4064	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2022-40664
CVE-2022-40152	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2022-40152
CVE-2022-40150	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2022-40150
CVE-2022-40149	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2022-40149
CVE-2022-40146	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2022-40146
CVE-2022-3786	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2022-37866
CVE-2022-3636 4	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2022-36364
CVE-2022-3510	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2022-3510
CVE-2022-3509	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2022-3509
CVE-2022-34169	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2022-34169
CVE-2022-32532	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2022-32532
CVE-2022-3171	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2022-3171

CVEID	Gravidade	CVEdetalhes URL
CVE-2022-25 647	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2022-25647
CVE-2022-24758	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2022-24758
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-2021-34 538	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-2021-3 684	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2021-31684
CVE-2021-28 165	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2021-28165
CVE-2021-26919	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2021-26919
CVE-2021-25 646	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2021-25646
CVE-2020-9492	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2020-9492

CVEID	Gravidade	CVEdetalhes URL
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-2020-3618	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
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

CVEID	Gravidade	CVEdetalhes URL
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
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

CVEID	Gravidade	CVEdetalhes URL
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
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

CVEID	Gravidade	CVEdetalhes URL
CVE-2020-10 672	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2020-10672
CVE-2020-10650	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2020-10650
CVE-2019-17 56	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2019-17566
CVE-2019-16 869	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2019-16869
CVE-2019-14 439	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2019-14439
CVE-2019-12 2086	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2019-12086
CVE-2019-10 172	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2019-10172
CVE-2019-1099	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2019-10099
CVE-2019-0205	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2019-0205
CVE-2019-0204	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2019-0204
CVE-2018-8 768	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

CVEID	Gravidade	CVEdetalhes URL
CVE-2018-3 827	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2018-3827
CVE-2018-3258	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2018-3258
CVE-2018-17 190	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2018-17190
CVE-2018-12 023	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2018-12023
CVE- 2018-12022	HIGH	https://nvd.nist.gov/vuln/detail/CVE - 2018-12022
CVE-2018-18 04	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2018-11804
CVE-2018-17 93	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2018-11793
CVE-2018-10 936	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2018-10936
CVE-2018-10 054	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2018-10054
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-2017-7 687	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2017-7687
CVE-2017-7656	HIGH	https://nvd.nist.gov/vuln/detail/CVE -2017-7656

CVEID	Gravidade	CVEdetalhes URL
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-2017-1528	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
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

CVEID	Gravidade	CVEdetalhes URL
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-2012-08 81	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2012-0881

7.2.0 corrigiu vulnerabilidades e exposições comuns para mecanismos principais

A tabela a seguir lista todas as correções CVEs na Amazon EMR 7.2.0.

CVEID	Gravidade	CVEdetalhes URL
CVE-2019-12 402	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2019-12402
CVE-2021-41561	HIGH	https://nvd.nist.gov/vuln/detail/CVE-2021-41561

Componentes da versão 7.2.0

O Amazon EMR 7.2.0 oferece suporte aos seguintes aplicativos:

Motores principais:

AmazonCloudWatchAgent, Delta, Flink, Hadoop,, HiveHBase, HudiHCatalog, Livy, Iceberg, Phoenix, Spark, Tez, Trino JupyterEnterpriseGateway, Zookeeper

Extras:

Hue, Oozie JupyterHubMXNet, Pig, Presto, Sqoop, Zepelin TensorFlow

Para obter mais detalhes sobre os pacotes suportados, consulte os [detalhes do componente da versão 7.2.0](#). Para obter mais informações sobre os principais mecanismos e extras, consulte [Amazon EMR Standard Support](#).

Detalhes dos componentes da versão 7.2.0

Consulte a tabela a seguir para obter mais informações sobre os pacotes Extras na Amazon EMR 7.2.0.

Aplicativo	Tipo de aplicativo	Pacote	Tipo de pacote	Versão do pacote
AmazonCloudWatchAgent	Serviços	hadoop-hdfs-journalnode	Extras	3.3.6.amzn.4-1.amzn2023
AmazonCloudWatchAgent	Serviços	ranger-kms	Extras	2.0.0-1.amzn2023
AmazonCloudWatchAgent	Serviços	s3-dist-cp	Extras	2.32.0-1.amzn2023
AmazonCloudWatchAgent	Serviços	tratador do zoológico	Extras	3.9.1.amzn.1-1.amzn2023
AmazonCloudWatchAgent	Serviços	zookeeper-server	Extras	3.9.1.amzn.1-1.amzn2023

Aplicativo	Tipo de aplicativo	Pacote	Tipo de pacote	Versão do pacote
Flink	Serviços	hadoop-client	Extras	3.3.6.amzn.4-1.amzn2023
Flink	Serviços	hadoop-hdfs	Extras	3.3.6.amzn.4-1.amzn2023
Flink	Serviços	hadoop-hdfs-datanode	Extras	3.3.6.amzn.4-1.amzn2023
Flink	Serviços	hadoop-hdfs-journalnode	Extras	3.3.6.amzn.4-1.amzn2023
Flink	Serviços	hadoop-hdfs-namenode	Extras	3.3.6.amzn.4-1.amzn2023
Flink	Serviços	hadoop-httfs	Extras	3.3.6.amzn.4-1.amzn2023
Flink	Serviços	hadoop-kms	Extras	3.3.6.amzn.4-1.amzn2023
Flink	Serviços	hadoop-mapreduce	Extras	3.3.6.amzn.4-1.amzn2023
Flink	Serviços	hadoop-mapreduce-historyserver	Extras	3.3.6.amzn.4-1.amzn2023
Flink	Serviços	hadoop	Extras	3.3.6.amzn.4-1.amzn2023
Flink	Serviços	hadoop-yarn-nodemanager	Extras	3.3.6.amzn.4-1.amzn2023
Flink	Serviços	hadoop-yarn-proxyserver	Extras	3.3.6.amzn.4-1.amzn2023

Aplicativo	Tipo de aplicativo	Pacote	Tipo de pacote	Versão do pacote
Flink	Serviços	hadoop-yarn-resourcemanager	Extras	3.3.6.amzn.4-1.amzn2023
Flink	Serviços	hadoop-yarn-timelineserver	Extras	3.3.6.amzn.4-1.amzn2023
Flink	Serviços	hudi	Extras	0.14.1.amzn.1-1.amzn2023
Flink	Serviços	ranger-kms	Extras	2.0.0-1.amzn2023
Flink	Serviços	hadoop	Extras	3.3.6.amzn.4-1.amzn2023
HBase	Serviços	s3-dist-cp	Extras	2.32.0-1.amzn2023
HCatalog	Serviços	hadoop-yarn-proxyserver	Extras	3.3.6.amzn.4-1.amzn2023
HCatalog	Serviços	mariadb-connector-java	Extras	2.7.2-1
Hive	Serviços	hadoop-httfs	Extras	3.3.6.amzn.4-1.amzn2023
Hive	Serviços	hadoop-yarn-proxyserver	Extras	3.3.6.amzn.4-1.amzn2023
Hive	Serviços	hive-hbase	Extras	3.1.3.amzn.11-1.amzn2023

Aplicativo	Tipo de aplicativo	Pacote	Tipo de pacote	Versão do pacote
Hive	Serviços	mariadb-connector-java	Extras	2.7.2-1
Hive	Serviços	s3-dist-cp	Extras	2.32.0-1.amzn2023
Phoenix	Serviços	s3-dist-cp	Extras	2.32.0-1.amzn2023
Spark	Serviços	núcleo de dados de ignição	Extras	3.5.1.amzn.0-1.amzn2023
Tez	Serviços	hadoop-mapreduce	Extras	3.3.6.amzn.4-1.amzn2023
Tez	Serviços	hadoop-mapreduce-historyserver	Extras	3.3.6.amzn.4-1.amzn2023
Tez	Serviços	hadoop-yarn-proxyserver	Extras	3.3.6.amzn.4-1.amzn2023
Trino	Serviços	grandes utilitários	Extras	1.2.0-1.amzn2023
Trino	Serviços	hadoop-client	Extras	3.3.6.amzn.4-1.amzn2023
Trino	Serviços	hadoop-hdfs	Extras	3.3.6.amzn.4-1.amzn2023
Trino	Serviços	hadoop-hdfs-datanode	Extras	3.3.6.amzn.4-1.amzn2023
Trino	Serviços	hadoop-hdfs-journalnode	Extras	3.3.6.amzn.4-1.amzn2023

Aplicativo	Tipo de aplicativo	Pacote	Tipo de pacote	Versão do pacote
Trino	Serviços	hadoop-hdfs-namenode	Extras	3.3.6.amzn.4-1.amzn2023
Trino	Serviços	hadoop-kms	Extras	3.3.6.amzn.4-1.amzn2023
Trino	Serviços	hadoop-mapreduce	Extras	3.3.6.amzn.4-1.amzn2023
Trino	Serviços	hadoop-mapreduce-historyserver	Extras	3.3.6.amzn.4-1.amzn2023
Trino	Serviços	hadoop	Extras	3.3.6.amzn.4-1.amzn2023
Trino	Serviços	hadoop-yarn-nodemanager	Extras	3.3.6.amzn.4-1.amzn2023
Trino	Serviços	hadoop-yarn-proxyserver	Extras	3.3.6.amzn.4-1.amzn2023
Trino	Serviços	hadoop-yarn-resourcemanager	Extras	3.3.6.amzn.4-1.amzn2023
Trino	Serviços	hadoop-yarn-timeline-server	Extras	3.3.6.amzn.4-1.amzn2023
Trino	Serviços	hive	Extras	3.1.3.amzn.11-1.amzn2023
Trino	Serviços	hive-hcatalog-server	Extras	3.1.3.amzn.11-1.amzn2023

Aplicativo	Tipo de aplicativo	Pacote	Tipo de pacote	Versão do pacote
Trino	Serviços	mariadb-connector-java	Extras	2.7.2-1
Trino	Serviços	ranger-kms	Extras	2.0.0-1.amzn2023
Trino	Serviços	tratador do zoológico	Extras	3.9.1.amzn.1-1.amzn2023
Trino	Serviços	zookeeper-server	Extras	3.9.1.amzn.1-1.amzn2023
Trino	Serviços	bigtop-groovy	Extras	2.5.4-1.amzn2023
Trino	Serviços	bigtop-jsvc	Extras	1.2.4-1.amzn2023
Trino	Serviços	hadoop	Extras	3.3.6.amzn.4-1.amzn2023
Trino	Serviços	hadoop-lzo	Extras	0.4.19-1.amzn2023
Trino	Serviços	catálogo hive-h	Extras	3.1.3.amzn.11-1.amzn2023
Trino	Serviços	hive-jdbc	Extras	3.1.3.amzn.11-1.amzn2023
Zookeeper	Serviços	emrfs	Extras	2.63.0-1.amzn2023

Aplicativo	Tipo de aplicativo	Pacote	Tipo de pacote	Versão do pacote
Zookeeper	Serviços	hadoop-client	Extras	3.3.6.amzn.4-1.amzn2023
Zookeeper	Serviços	hadoop-hdfs	Extras	3.3.6.amzn.4-1.amzn2023
Zookeeper	Serviços	hadoop-hdfs-datanode	Extras	3.3.6.amzn.4-1.amzn2023
Zookeeper	Serviços	hadoop-hdfs-journalnode	Extras	3.3.6.amzn.4-1.amzn2023
Zookeeper	Serviços	hadoop-hdfs-namenode	Extras	3.3.6.amzn.4-1.amzn2023
Zookeeper	Serviços	hadoop-https	Extras	3.3.6.amzn.4-1.amzn2023
Zookeeper	Serviços	hadoop-kms	Extras	3.3.6.amzn.4-1.amzn2023
Zookeeper	Serviços	hadoop-mapreduce	Extras	3.3.6.amzn.4-1.amzn2023
Zookeeper	Serviços	hadoop-mapreduce-historyserver	Extras	3.3.6.amzn.4-1.amzn2023
Zookeeper	Serviços	hadoop	Extras	3.3.6.amzn.4-1.amzn2023
Zookeeper	Serviços	hadoop-yarn-nodemanager	Extras	3.3.6.amzn.4-1.amzn2023
Zookeeper	Serviços	hadoop-yarn-proxyserver	Extras	3.3.6.amzn.4-1.amzn2023

Aplicativo	Tipo de aplicativo	Pacote	Tipo de pacote	Versão do pacote
Zookeeper	Serviços	hadoop-yarn-resource-manager	Extras	3.3.6.amzn.4-1.amzn2023
Zookeeper	Serviços	hadoop-yarn-timelineserver	Extras	3.3.6.amzn.4-1.amzn2023
Zookeeper	Serviços	ranger-kms	Extras	2.0.0-1.amzn2023

Versões do aplicativo 7.2.0

As seguintes aplicações são compatíveis com esta versão: [AmazonCloudWatchAgent](#), [Delta](#), [Flink](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-7.2.0	emr-7.1.0	emr-7.0.0
AWS SDK para 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

	emr-7.2.0	emr-7.1.0	emr-7.0.0
Scala	2.12.17	2.12.17	2.12.17
AmazonCloudWatchAgent	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-aman-0	1.4.3-amzn-0	q1.4.2-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
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

	emr-7.2.0	emr-7.1.0	emr-7.0.0
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	0.10.2	0.10.2	0.10.2
Trino (PrestoSQL)	436	435	426
Zeppelin	0.10.1	0.10.1	0.10.1
ZooKeeper	3.9.1	3.9.1	3.5.10

Notas de lançamento da versão 7.2.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 7.2.0 da Amazon. As alterações são relativas à versão 7.2.0.

Novos atributos

- Atualizações da aplicação: Amazon EMR 7.2.0 application upgrades include Iceberg 1.5.0-amzn-0 and Delta 3.1.0.
- EMRA Amazon adiciona suporte para que você possa usar outros aplicativos HBase, como Flink e Hive, com a classe de armazenamento Amazon S3 Express One Zone.
- Esta versão adiciona a capacidade de ler objetos restaurados, para que você possa ler objetos Glacier de um local do S3 com o S3A protocolo. Esse recurso funciona com Spark, Flink e Hive.
- Rótulos de nós com escalabilidade gerenciada - Use rótulos de nós com suas instâncias com base no tipo de mercado ou tipo de nó para melhorar a escalabilidade automática da Amazon. EMR Para obter mais informações, consulte [Usando escalabilidade gerenciada na Amazon EMR](#).

Problemas conhecidos

- O Python 3.11 não é compatível com o Studio. EMR

Alterações, melhorias e problemas resolvidos

- Esta versão corrige um problema de impasse que pode ocorrer durante as operações internas de limpeza de etapas. Essa operação gerencia o ciclo de vida das etapas à medida que elas são concluídas no EMR cluster. Esse problema afeta EMR operações críticas da Amazon, como operação por etapas e escalabilidade.
- Esta versão resolve um problema em que clusters personalizados com determinados arquivos de log preexistentes podem fazer com AMIs que o daemon de gerenciamento de EMR log da Amazon falhe.
- O Amazon EMR 7.2.0 atualiza o EMR daemon da Amazon responsável pelas atividades de gerenciamento e monitoramento de clusters da v1 para a v2. AWS SDK
- Quando você inicia um cluster com a versão de patch mais recente do Amazon EMR 5.36 ou superior, 6.6 ou superior ou 7.0 ou superior, a Amazon EMR usa a versão mais recente do Amazon Linux 2023 ou Amazon Linux 2 para a Amazon padrão. EMR AMI Para obter mais informações, consulte [Usando o Amazon Linux padrão AMI para a Amazon EMR](#).

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2023.5.2 240708.0	6.1.96-102.177.amzn2023	8 de julho de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central), Israel (Tel Aviv), Oeste do Canadá (Calgary), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia)

Versões Java padrão 7.2.0

A EMR versão 7.2 e superior da Amazon vem com o Amazon Corretto 17 (construído no JDK Open) por padrão para aplicativos compatíveis com o Corretto 17 JDK (17), com exceção do Apache Livy.

A tabela a seguir mostra as versões padrão do Java para aplicativos na Amazon EMR 7.2.0. Se você quiser alterar o padrão JVM em seu cluster, siga as [Configurar aplicações para usar uma Máquina](#)

[Virtual Java específica](#) instruções em cada aplicativo executado no cluster. Você só pode usar uma versão de runtime do Java para um cluster. A Amazon EMR não oferece suporte à execução de diferentes nós ou aplicativos em diferentes versões de tempo de execução no mesmo cluster.

Aplicativo	Versão Java/Amazon Corretto (o padrão é negrito)
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
Faísca RAPIDS	17, 11, 8
Sqoop	8
Tez	17, 11, 8
Trino	17

Aplicativo	Versão Java/Amazon Corretto (o padrão é negrito)
Zeppelin	8
Pig	8
Zookeeper	8

Versões do componente 7.2.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>adot-java-agent</code>	1.31.0	Um Java Agent que coleta métricas de daemons de aplicações.
<code>delta</code>	3.1.0	O Delta Lake é um formato de tabela aberto para grandes conjuntos de dados analíticos
<code>emr-amazon-cloudwatch-agent</code>	1.300032.2-amzn-0	Um aplicativo que coleta métricas internas em nível de sistema e métricas de aplicativ

Componente	Version (Versão)	Descrição
		os personalizadas de instâncias da Amazon. EC2
emr-ddb	5.4.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	3.11.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.15.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-notebook-env	1.16.0	Ambiente Conda para bloco de anotações do emr, que inclui o jupyter enterprise gateway
emr-s3-dist-cp	2.32.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	2.11.0	EMRConector S3Select
emr-wal-cli	1.2.2	Cli usada para lista/exclusão de emrwal.
emrfs	2.63.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.18.1-amzn-1	Scripts do cliente da linha de comando e aplicativos do Apache Flink.

Componente	Version (Versão)	Descrição
flink-jobmanager-config	1.18.1-amzn-1	Gerenciando recursos em EMR nós para o Apache JobManager Flink.
hadoop-client	3.3.6-aman-4	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	3.3.6-aman-4	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	3.3.6-aman-4	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	3.3.6-aman-4	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	3.3.6-aman-4	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httpfs-server	3.3.6-aman-4	HTTPendpoint para HDFS operações.
hadoop-kms-server	3.3.6-aman-4	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	3.3.6-aman-4	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	3.3.6-aman-4	YARNserviço para gerenciar contêineres em um nó individual.

Componente	Version (Versão)	Descrição
hadoop-yarn-resourcemanager	3.3.6-aman-4	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	3.3.6-aman-4	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	2.4.17-amzn-6	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	2.4.17-amzn-6	Serviço para atender uma ou mais HBase regiões.
hbase-client	2.4.17-amzn-6	HBasecliente de linha de comando.
hbase-rest-server	2.4.17-amzn-6	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	2.4.17-amzn-6	Serviço que fornece um endpoint Thrift para. HBase
hbase-operator-tools	2.4.17-amzn-6	Ferramenta de reparo para HBase clusters Apache.
hcatalog-client	3.1.3-amzn-11	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.

Componente	Version (Versão)	Descrição
hcatalog-server	3.1.3-amzn-11	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	3.1.3-amzn-11	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	3.1.3-amzn-11	O cliente da linha de comando do Hive.
hive-hbase	3.1.3-amzn-11	Cliente do Hive-hbase.
hive-metastore-server	3.1.3-amzn-11	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	3.1.3-amzn-11	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.14.1-amzn-1	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.
hudi-presto	0.14.1-amzn-1	Biblioteca de pacotes para executar o Presto com o Hudi.
hudi-trino	0.14.1-amzn-1	Biblioteca de pacotes para execução do Presto com o Hudi.

Componente	Version (Versão)	Descrição
hudi-spark	0.14.1-amzn-1	Biblioteca de pacotes para executar o Spark com o Hudi.
hue-server	4.11.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
iceberg	1.5.0-aman-0	Apache Iceberg é um formato de tabela aberta para conjuntos de dados analíticos imensos
jupyterhub	1.5.0	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.8.0 - incubação	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mxnet	1.9.1	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mariadb-server	5.5.68+	Servidor de banco de dados MariaDB.
nvidia-cuda	11.8.0	Drivers NVIDIA e toolkit CUDA
oozie-client	5.2.1	O cliente da linha de comando do Oozie.
oozie-server	5.2.1	O serviço que aceita solicitações de fluxo de trabalho do Oozie.

Componente	Version (Versão)	Descrição
opencv	4.7.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	5.1.3	As bibliotecas do phoenix para servidor e cliente
phoenix-connectors	5.1.3	Conectores do Apache Phoenix para Spark-3
phoenix-query-server	6.0.0	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0,285-amzn-0	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0,285-amzn-0	O serviço que executa partes de uma consulta.
presto-client	0,285-amzn-0	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
trino-coordinator	436-amzn-0	Serviço que aceita consultas e gerencia a execução de consultas entre trino-workers.
trino-worker	436-amzn-0	O serviço que executa partes de uma consulta.

Componente	Version (Versão)	Descrição
trino-client	436-amzn-0	Cliente de linha de comando do Trino instalado em mestres de espera de um cluster de alta disponibilidade (HA) em que o servidor do Trino não é iniciado.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	4.3.2	O projeto R para computação estatística
ranger-kms-server	2.0.0	Sistema de gerenciamento de chaves Apache Ranger
spark-client	3.5.1-amzn-0	Os clientes da linha de comando do Spark.
spark-history-server	3.5.1-amzn-0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	3.5.1-amzn-0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	3.5.1-amzn-0	Bibliotecas do Apache Spark necessárias para escravos. YARN
spark-rapids	24.02.0-aman-0	RAPIDSPlugin Nvidia Spark que acelera o Apache Spark com o. GPUs

Componente	Version (Versão)	Descrição
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	2.11.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.10.2-amzn-9	O YARN aplicativo e as bibliotecas tez.
tez-on-worker	0.10.2-amzn-9	O YARN aplicativo tez e as bibliotecas para nós de trabalho.
zeppelin-server	0.10.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.9.1-amzn-1	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.9.1-amzn-1	ZooKeeper cliente de linha de comando.

Classificações de configuração 7.2.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para obter mais informações, consulte [Configurar aplicações](#).

Ocorrem ações de reconfiguração quando você especifica uma configuração para grupos de instâncias em um cluster em execução. A Amazon EMR só inicia ações de reconfiguração para as classificações que você modifica. Para obter mais informações, consulte [Reconfigurar um grupo de instâncias de um cluster em execução](#).

classificações do emr-7.2.0

Classificações	Descrição	Ações de reconfiguração
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.	Restarts the ResourceManager service.
container-executor	Altere os valores no arquivo container-executor.cfg YARN do Hadoop.	Not available.
container-log4j	Altere os valores no arquivo container-log4j.properties YARN do Hadoop.	Not available.
core-site	Alterar os valores no arquivo core-site.xml do 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	Alterar as configurações relacionadas ao docker.	Not available.

Classificações	Descrição	Ações de reconfiguração
emrfs-site	Altere EMRFS as configurações.	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	Alterar as configurações do flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Alterar as configurações de log4j.properties no Flink.	Restarts Flink history server.
flink-log4j-session	Alterar as configurações de log4j-yarn-session.properties no Flink para sessão do Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.	Restarts Flink history server.

Classificações	Descrição	Ações de reconfiguração
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do 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	Alterar os valores no arquivo log4j.properties do 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	Alterar a configuração do servidor SSL no Hadoop	Not available.
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop	Not available.

Classificações	Descrição	Ações de reconfiguração
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Mude os valores no HBase ambiente.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.	Not available.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.	This classification should not be reconfigured.
hdfs-env	Mude os valores no HDFS ambiente.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classificações	Descrição	Ações de reconfiguração
hdfs-site	Altere os valores no HDFS hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Mude os valores no HCatalog ambiente.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Configurações EMR selecionadas pela Amazon para o Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.	Not available.

Classificações	Descrição	Ações de reconfiguração
hive-parquet-logging	Alterar os valores no arquivo <code>parquet-logging.properties</code> do Hive.	Not available.
hive-env	Alterar os valores no ambiente do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive <code>schemaTool CLI</code> commands to verify <code>hive-metastore</code> .
hive-exec-log4j2	Altere os valores no arquivo <code>hive-exec-log 4j2.properties</code> do Hive.	Not available.
hive-llap-daemon-log4j2	Altere os valores no arquivo <code>llap-daemon-log 4j2.properties</code> do Hive.	Not available.
hive-log4j2	Alterar os valores no arquivo <code>hive-log4j2.properties</code> do Hive.	Not available.
hive-site	Alterar os valores no arquivo <code>hive-site.xml</code> do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive <code>schemaTool CLI</code> commands to verify <code>hive-metastore</code> . Also restarts Oozie and Zeppelin.
hiveserver2-site	Alterar os valores no arquivo <code>hiveserver2-site.xml</code> do Hive Server2.	Not available.
hue-ini	Alterar os valores no arquivo <code>ini</code> do Hue	Restarts Hue. Also activates Hue <code>config override CLI</code> commands to pick up new configurations.

Classificações	Descrição	Ações de reconfiguração
httpfs-env	Mude os valores no HTTPFS ambiente.	Restarts Hadoop Httpfs service.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.	Not available.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.	Restarts Hadoop-KMS service.
hadoop-kms-java-home	Mude a página inicial do Java do KMS Hadoop	Not available.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.	Not available.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Altere os valores no ambiente do Hudi.	Not available.
hudi-defaults	Alterar os valores no arquivo hudi-defaults.conf do Hudi.	Not available.
iceberg-defaults	Alterar os valores no arquivo iceberg-defaults.conf do Iceberg.	Not available.
delta-defaults	Alterar os valores no arquivo delta-defaults.conf do Delta.	Not available.

Classificações	Descrição	Ações de reconfiguração
jupyter-notebook-conf	Alterar os valores no arquivo <code>jupyter_notebook_config.py</code> do Notebook Jupyter.	Not available.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo <code>jupyterhub_config.py</code> do.	Not available.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.	Not available.
jupyter-sparkmagic-conf	Altere os valores no arquivo <code>config.json</code> do Sparkmagic.	Not available.
livy-conf	Alterar os valores no arquivo <code>livy.conf</code> do Livy.	Restarts Livy Server.
livy-env	Alterar os valores no ambiente do Livy.	Restarts Livy Server.
livy-log4j2	Alterar as configurações de <code>log4j2.properties</code> do Livy.	Restarts Livy Server.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Altere os valores no arquivo <code>mapred-site.xml</code> do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Alterar os valores no ambiente do Oozie.	Restarts Oozie.
oozie-log4j	Alterar os valores no arquivo <code>oozie-log4j.properties</code> do Oozie.	Restarts Oozie.

Classificações	Descrição	Ações de reconfiguração
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.	Not available.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.	Not available.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.	Not available.
pig-env	Alterar os valores no ambiente do Pig.	Not available.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.	Restarts Oozie.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.	Not available.
presto-log	Alterar os valores no arquivo log.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Alterar os valores no arquivo config.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-env	Altere os valores no arquivo <code>presto-env.sh</code> do Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Altere os valores no arquivo <code>node.properties</code> do Presto.	Not available.
presto-connector-blackhole	Alterar os valores no arquivo <code>blackhole.properties</code> do Presto.	Not available.
presto-connector-cassandra	Alterar os valores no arquivo <code>cassandra.properties</code> do Presto.	Not available.
presto-connector-hive	Alterar os valores no arquivo <code>hive.properties</code> do Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Alterar os valores no arquivo <code>jmx.properties</code> do Presto.	Not available.
presto-connector-kafka	Alterar os valores no arquivo <code>kafka.properties</code> do Presto.	Not available.
presto-connector-lakeformat ion	Alterar os valores no arquivo <code>lakeformation.properties</code> do Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Alterar os valores no arquivo <code>localfile.properties</code> do Presto.	Not available.
presto-connector-memory	Alterar os valores no arquivo <code>memory.properties</code> do Presto.	Not available.
presto-connector-mongodb	Alterar os valores no arquivo <code>mongodb.properties</code> do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.	Not available.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.	Not available.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.	Not available.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.	Not available.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.	Not available.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.	Not available.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.	Not available.
trino-log	Alterar os valores no arquivo log.properties do Trino.	Restarts Trino-Server (for Trino)
trino-config	Alterar os valores no arquivo config.properties do Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Alterar os valores no arquivo password-authenticator.properties do Trino.	Restarts Trino-Server (for Trino)
trino-env	Alterar os valores no arquivo trino-env.sh do Trino.	Restarts Trino-Server (for Trino)

Classificações	Descrição	Ações de reconfiguração
trino-node	Alterar os valores no arquivo <code>node.properties</code> do Trino.	Not available.
trino-connector-blackhole	Alterar os valores no arquivo <code>blackhole.properties</code> do Trino.	Not available.
trino-connector-cassandra	Alterar os valores no arquivo <code>cassandra.properties</code> do Trino.	Not available.
trino-connector-delta	Alterar os valores no arquivo <code>delta.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Alterar os valores no arquivo <code>hive.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Alterar os valores no arquivo <code>exchange-manager.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Alterar os valores no arquivo <code>iceberg.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Alterar os valores no arquivo <code>hudi.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Alterar os valores no arquivo <code>jmx.properties</code> do Trino.	Not available.
trino-connector-kafka	Alterar os valores no arquivo <code>kafka.properties</code> do Trino.	Not available.
trino-connector-localfile	Alterar os valores no arquivo <code>localfile.properties</code> do Trino.	Not available.
trino-connector-memory	Alterar os valores no arquivo <code>memory.properties</code> do Trino.	Not available.

Classificações	Descrição	Ações de reconfiguração
trino-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Trino.	Not available.
trino-connector-mysql	Alterar os valores no arquivo mysql.properties do Trino.	Not available.
trino-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Trino.	Not available.
trino-connector-raptor	Alterar os valores no arquivo raptor.properties do Trino.	Not available.
trino-connector-redis	Alterar os valores no arquivo redis.properties do Trino.	Not available.
trino-connector-redshift	Alterar os valores no arquivo redshift.properties do Trino.	Not available.
trino-connector-tpch	Alterar os valores no arquivo tpch.properties do Trino.	Not available.
trino-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Trino.	Not available.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.	Restarts Ranger KMS Server.
ranger-kms-logback	Altere os valores no arquivo kms-logback.xml do Ranger. KMS	Not available.

Classificações	Descrição	Ações de reconfiguração
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.	Not available.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Alterar os valores no ambiente do Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.	Not available.
spark-log4j2	Alterar os valores no arquivo log4j2.properties do Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Alterar os valores no ambiente do Sqoop.	Not available.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.	Not available.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.	Not available.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.	Restart Oozie and HiveServer2.

Classificações	Descrição	Ações de reconfiguração
yarn-env	Mude os valores no YARN ambiente.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.	Restarts Zeppelin.
zeppelin-site	Alterar as definições da configuração em zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.	Restarts Zookeeper server.
zookeeper-logback	Altere os valores no ZooKeeper arquivo logback.xml do.	Restarts Zookeeper server.
emr-metrics	Altere as configurações da métrica emr para esse nó.	Restarts the CloudWatchAgent service.

Registro de alterações 7.2.0

Registro de alterações da versão 7.2.0 e das notas de versão

Data	Evento	Descrição
2024-07-25	Publicação de documentos	Notas de lançamento do Amazon EMR 7.2.0 publicadas pela primeira vez
2024-07-16	Lançamento inicial	Amazon EMR 7.2.0 foi implantado pela primeira vez nas regiões comerciais iniciais

Amazon EMR versão 7.1.0

Versões do aplicativo 7.1.0

As seguintes aplicações são compatíveis com esta versão: [AmazonCloudWatchAgent](#), [Delta](#), [Flink](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-7.2.0	emr-7.1.0	emr-7.0.0
AWS SDK para 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	2.12.17	2.12.17	2.12.17
Amazon CloudWatch Agent	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-aman-0	1.4.3-amzn-0	q1.4.2-amzn-0
Jupyter Enterprise Gateway	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	-	-	-

	emr-7.2.0	emr-7.1.0	emr-7.0.0
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	0.10.2	0.10.2	0.10.2
Trino (PrestoSQL)	436	435	426
Zeppelin	0.10.1	0.10.1	0.10.1
ZooKeeper	3.9.1	3.9.1	3.5.10

Notas de lançamento do 7.1.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 7.1.0 da Amazon. As alterações são relativas à 7.1.0.

Novos atributos

- Atualizações da aplicação: Amazon EMR 7.1.0 application upgrades include Livy 0.8.0, Trino 435, and ZooKeeper 3.9.1.
- [Substituição de nós não íntegros](#) — Com o Amazon EMR 7.1.0 e versões posteriores, a substituição de nós não íntegros é ativada por padrão, então a Amazon EMR substituirá seus nós não íntegros. Para evitar afetar seus fluxos de trabalho existentes nas EMR versões 7.0.0 e anteriores da Amazon, a substituição não íntegra de nós será desativada se você habilitar a proteção contra encerramento em seu cluster.

- [CloudWatch Agente](#) — Configure o CloudWatch agente para usar métricas adicionais do sistema, adicionar métricas de aplicativos e alterar o destino das métricas com a EMR configuração da AmazonAPI.

Problemas conhecidos

- O Python 3.11 não é compatível com o Studio. EMR

Alterações, melhorias e problemas resolvidos

- Enquanto o Amazon EMR 7.1.0 oferece suporte ao Python 3.9 por padrão, o Livy 0.8.0 e o Spark no EMR Amazon 7.1.0 oferecem suporte ao Python 3.11.
- Esta versão corrige o problema da necessidade de executar cada linha, uma por vez, ao usar PySpark com o Python versão 3.11.
- Atualização do Zeppelin — O Amazon EMR 7.1.0 inclui uma atualização do Zeppelin para Java v2. AWS SDK Essa atualização permite que um notebook Zeppelin S3 aceite um fornecedor de materiais de criptografia personalizados. O AWS SDK for Java v2 remove a EncryptionMaterialsProvider interface. Ao atualizar para o Amazon EMR 7.1.0, você deve implementar a [interface Keyring](#) se quiser usar criptografia personalizada. Para ver um exemplo de como implementar a interface Keyring, consulte [KmsKeyring.java](#).
- Ao atualizar para a EMR versão 7.1.0 da Amazon, altere seu provedor de chaves personalizadas para criptografia de disco local para gerar chaves usando o AES algoritmo deAES//GCM. NoPadding Se você não atualizar o algoritmo, a criação do cluster poderá falhar com o erro. Local disk encryption failed on master instance (i-123456789) due to internal error. Para obter mais informações sobre como criar um provedor de chave personalizada, consulte [Criação de um provedor de chave personalizada](#).
- O Amazon EMR 7.1.0 melhora a resiliência de um nó em condições de pouco espaço em disco, melhorando a lógica de truncamento de log para arquivos com identificadores de arquivo abertos.
- Esta versão aprimora a lógica de codificação e decodificação para minimizar o risco de corrupção de dados e falha do nó com os EMR daemons da Amazon que leem e gravam arquivos ao reiniciar um nó.
- Quando você inicia um cluster com a versão de patch mais recente do Amazon EMR 5.36 ou superior, 6.6 ou superior ou 7.0 ou superior, a Amazon EMR usa a versão mais recente do

Amazon Linux 2023 ou Amazon Linux 2 para a Amazon padrão. EMR AMI Para obter mais informações, consulte [Usando o Amazon Linux padrão AMI para a Amazon EMR](#).

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2023.5.2 240708.0	6.1.96-102.177.amzn2023	23 de julho de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			China (Ningxia), Ásia Pacífico (Hyderabad), Oriente Médio (), Europa (EspanhaUAE), Europa (Zurique), Ásia-Pacífico (Melbourne), Israel (Tel Aviv), Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2023.3.2 240219.0	6.1.77-99.164.amzn2023	8 de maio de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), Oeste do Canadá (Calgary), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia)

Versões Java padrão 7.1.0

A EMR versão 7.1 e superior da Amazon vem com o Amazon Corretto 17 (construído no JDK Open) por padrão para aplicativos compatíveis com o Corretto 17 JDK (17), com exceção do Apache Livy.

A tabela a seguir mostra as versões padrão do Java para aplicativos na Amazon EMR 7.1.0. Se você quiser alterar o padrão JVM em seu cluster, siga as [Configurar aplicações para usar uma Máquina Virtual Java específica](#) instruções em cada aplicativo executado no cluster. Você só pode usar uma versão de runtime do Java para um cluster. A Amazon EMR não oferece suporte à execução de diferentes nós ou aplicativos em diferentes versões de tempo de execução no mesmo cluster.

Aplicativo	Versão Java/Amazon Corretto (o padrão é negrito)
Delta	17, 11, 8
Flink	17, 11, 8
HBase	11, 8
HCatalog	17, 11, 8
Hadoop	17, 11, 8
Hive	17, 11, 8

Aplicativo	Versão Java/Amazon Corretto (o padrão é negrito)
Hudi	17, 11, 8
Iceberg	17, 11, 8
Livy	17, 11, 8
Oozie	17, 11, 8
Phoenix	8
PrestoDB	8
Spark	17, 11, 8
Faísca RAPIDS	17, 11, 8
Sqoop	8
Tez	17, 11, 8
Trino	17
Zeppelin	8
Pig	8
Zookeeper	8

Versões do componente 7.1.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado myapp-component com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como 2.2-amzn-2.

Componente	Version (Versão)	Descrição
adot-java-agent	1.31.0	Um Java Agent que coleta métricas de daemons de aplicações.
delta	3.0.0	O Delta Lake é um formato de tabela aberto para grandes conjuntos de dados analíticos
emr-amazon-cloudwatch-agent	1.300032.2-amzn-0	Um aplicativo que coleta métricas internas em nível de sistema e métricas de aplicativos personalizadas de instâncias da Amazon. EC2
emr-ddb	5.3.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	3.10.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.14.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-notebook-env	1.7.0	Ambiente Conda para bloco de anotações do emr, que inclui o jupyter enterprise gateway

Componente	Version (Versão)	Descrição
emr-s3-dist-cp	2.31.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	2.10.0	EMRConector S3Select
emr-wal-cli	1.2.2	Cli usada para lista/exclusão de emrwal.
emrfs	2.62.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.18.1-amzn-0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
flink-jobmanager-config	1.18.1-amzn-0	Gerenciando recursos em EMR nós para o Apache JobManager Flink.
hadoop-client	3.3.6-aman-3	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	3.3.6-aman-3	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	3.3.6-aman-3	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	3.3.6-aman-3	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	3.3.6-aman-3	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.

Componente	Version (Versão)	Descrição
hadoop-https-server	3.3.6-aman-3	HTTP endpoint para HDFS operações.
hadoop-kms-server	3.3.6-aman-3	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	3.3.6-aman-3	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	3.3.6-aman-3	YARN serviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	3.3.6-aman-3	YARN serviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	3.3.6-aman-3	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	2.4.17-amzn-5	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	2.4.17-amzn-5	Serviço para atender uma ou mais HBase regiões.
hbase-client	2.4.17-amzn-5	HBase cliente de linha de comando.

Componente	Version (Versão)	Descrição
hbase-rest-server	2.4.17-amzn-5	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	2.4.17-amzn-5	Serviço que fornece um endpoint Thrift para HBase
hbase-operator-tools	2.4.17-amzn-5	Ferramenta de reparo para HBase clusters Apache.
hcatalog-client	3.1.3-amzn-10	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	3.1.3-amzn-10	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	3.1.3-amzn-10	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	3.1.3-amzn-10	O cliente da linha de comando do Hive.
hive-hbase	3.1.3-amzn-10	Cliente do Hive-hbase.
hive-metastore-server	3.1.3-amzn-10	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL

Componente	Version (Versão)	Descrição
hive-server2	3.1.3-amzn-10	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.14.1-amzn-0	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.
hudi-presto	0.14.1-amzn-0	Biblioteca de pacotes para executar o Presto com o Hudi.
hudi-trino	0.14.1-amzn-0	Biblioteca de pacotes para execução do Presto com o Hudi.
hudi-spark	0.14.1-amzn-0	Biblioteca de pacotes para executar o Spark com o Hudi.
hue-server	4.11.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
iceberg	1.4.3-amzn-0	Apache Iceberg é um formato de tabela aberta para conjuntos de dados analíticos imensos
jupyterhub	1.5.0	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.8.0 - incubação	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso

Componente	Version (Versão)	Descrição
mxnet	1.9.1	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mariadb-server	5.5.68+	Servidor de banco de dados MariaDB.
nvidia-cuda	11.8.0	Drivers NVIDIA e toolkit CUDA
oozie-client	5.2.1	O cliente da linha de comando do Oozie.
oozie-server	5.2.1	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	4.7.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	5.1.3	As bibliotecas do phoenix para servidor e cliente
phoenix-connectors	5.1.3	Conectores do Apache Phoenix para Spark-3
phoenix-query-server	5.1.3	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0,284-amzn-0	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0,284-amzn-0	O serviço que executa partes de uma consulta.

Componente	Version (Versão)	Descrição
presto-client	0,284-amzn-0	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
trino-coordinator	435-amzn-0	Serviço que aceita consultas e gerencia a execução de consultas entre trino-workers.
trino-worker	435-amzn-0	O serviço que executa partes de uma consulta.
trino-client	435-amzn-0	Cliente de linha de comando do Trino instalado em mestres de espera de um cluster de alta disponibilidade (HA) em que o servidor do Trino não é iniciado.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	4.3.2	O projeto R para computação estatística
ranger-kms-server	2.0.0	Sistema de gerenciamento de chaves Apache Ranger
spark-client	3.5.0-amzn-1	Os clientes da linha de comando do Spark.

Componente	Version (Versão)	Descrição
spark-history-server	3.5.0-amzn-1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	3.5.0-amzn-1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	3.5.0-amzn-1	Bibliotecas do Apache Spark necessárias para escravos. YARN
spark-rapids	23.10.0-amzn-1	RAPIDSPlugin Nvidia Spark que acelera o Apache Spark com o. GPUs
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	2.11.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.10.2-amzn-8	O YARN aplicativo e as bibliotecas tez.
tez-on-worker	0.10.2-amzn-8	O YARN aplicativo tez e as bibliotecas para nós de trabalho.
zeppelin-server	0.10.1	O notebook baseado na web que permite a análise de dados interativa.

Componente	Version (Versão)	Descrição
zookeeper-server	3.9.1-amzn-0	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.9.1-amzn-0	ZooKeeper cliente de linha de comando.

Classificações de configuração 7.1.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para obter mais informações, consulte [Configurar aplicações](#).

classificações do emr-7.1.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
container-executor	Altere os valores no arquivo <code>container-executor.cfg</code> YARN do Hadoop.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
docker-conf	Alterar as configurações relacionadas ao docker.
emrfs-site	Altere EMRFS as configurações.

Classificações	Descrição
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-session	Alterar as configurações de log4j-yarn-session.properties no Flink para sessão do Kubernetes/Yarn.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.

Classificações	Descrição
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-env	Mude os valores no HDFS ambiente.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive	Configurações EMR selecionadas pela Amazon para o Apache Hive.
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.

Classificações	Descrição
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-java-home	Mude a página inicial do Java do KMS Hadoop
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
hudi-env	Altere os valores no ambiente do Hudi.
hudi-defaults	Alterar os valores no arquivo hudi-defaults.conf do Hudi.

Classificações	Descrição
iceberg-defaults	Alterar os valores no arquivo iceberg-defaults.conf do Iceberg.
delta-defaults	Alterar os valores no arquivo delta-defaults.conf do Delta.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j2	Alterar as configurações de log4j2.properties do Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.

Classificações	Descrição
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.

Classificações	Descrição
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-lakeformation	Alterar os valores no arquivo lakeformation.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.

Classificações	Descrição
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.
trino-log	Alterar os valores no arquivo log.properties do Trino.
trino-config	Alterar os valores no arquivo config.properties do Trino.
trino-password-authenticator	Alterar os valores no arquivo password-authenticator.properties do Trino.
trino-env	Alterar os valores no arquivo trino-env.sh do Trino.
trino-node	Alterar os valores no arquivo node.properties do Trino.
trino-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Trino.
trino-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Trino.
trino-connector-delta	Alterar os valores no arquivo delta.properties do Trino.
trino-connector-hive	Alterar os valores no arquivo hive.properties do Trino.
trino-exchange-manager	Alterar os valores no arquivo exchange-manager.properties do Trino.
trino-connector-iceberg	Alterar os valores no arquivo iceberg.properties do Trino.

Classificações	Descrição
trino-connector-hudi	Alterar os valores no arquivo hudi.properties do Trino.
trino-connector-jmx	Alterar os valores no arquivo jmx.properties do Trino.
trino-connector-kafka	Alterar os valores no arquivo kafka.properties do Trino.
trino-connector-localfile	Alterar os valores no arquivo localfile.properties do Trino.
trino-connector-memory	Alterar os valores no arquivo memory.properties do Trino.
trino-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Trino.
trino-connector-mysql	Alterar os valores no arquivo mysql.properties do Trino.
trino-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Trino.
trino-connector-raptor	Alterar os valores no arquivo raptor.properties do Trino.
trino-connector-redis	Alterar os valores no arquivo redis.properties do Trino.
trino-connector-redshift	Alterar os valores no arquivo redshift.properties do Trino.
trino-connector-tpch	Alterar os valores no arquivo tpch.properties do Trino.
trino-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Trino.

Classificações	Descrição
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.
ranger-kms-logback	Altere os valores no arquivo kms-logback.xml do Ranger. KMS
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j2	Alterar os valores no arquivo log4j2.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.

Classificações	Descrição
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zeppelin-site	Alterar as definições da configuração em zeppelin-site.xml.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-logback	Altere os valores no ZooKeeper arquivo logback.xml do.
emr-metrics	Altere as configurações da métrica emr para esse nó.

Registro de alterações 7.1.0

Registro de alterações da versão 7.1.0 e das notas de lançamento

Data	Evento	Descrição
2024-05-10	Publicação de documentos	Notas de lançamento do Amazon EMR 7.1.0 publicadas pela primeira vez
2024-05-08	Implantação concluída	Amazon EMR 7.1.0 totalmente implantado em todas as regiões suportadas

Data	Evento	Descrição
2024-04-23	Lançamento inicial	Amazon EMR 7.1.0 foi implantado pela primeira vez nas regiões comerciais iniciais

Amazon EMR versão 7.0.0

Versões 7.0.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [AmazonCloudWatchAgent](#), [Delta](#), [Flink](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-7.2.0	emr-7.1.0	emr-7.0.0
AWS SDKpara 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	2.12.17	2.12.17	2.12.17
AmazonClo udWatchAgent	1.300032.2	1.300032.2	1.300031.1

	emr-7.2.0	emr-7.1.0	emr-7.0.0
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-aman-0	1.4.3-amzn-0	q1.4.2-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

	emr-7.2.0	emr-7.1.0	emr-7.0.0
TensorFlow	2.11.0	2.11.0	2.11.0
Tez	0.10.2	0.10.2	0.10.2
Trino (PrestoSQL)	436	435	426
Zeppelin	0.10.1	0.10.1	0.10.1
ZooKeeper	3.9.1	3.9.1	3.5.10

Notas da versão 7.0.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 7.0.0 da Amazon. As alterações são referentes à versão 6.15.0.

Novos atributos

- **Atualizações da aplicação:** Amazon EMR 7.0.0 application upgrades include Python 3.9, Spark 3.5, Flink 1.18, and Delta 3.0. Essa versão também adiciona suporte à aplicação do [CloudWatch Agente da Amazon](#) e remove o suporte ao Ganglia.
- **Amazon Corretto 17** — As EMR versões 7.0 e superiores da Amazon são fornecidas com o Amazon Corretto 17 (construído no OpenJDK) por padrão para aplicativos compatíveis com o Corretto 17 (JDK17), com exceção do Apache Livy. Para obter mais informações sobre as JDK versões suportadas dos aplicativos nesta versão, consulte [Versões padrão do Java 7.0.0](#).
- **Amazon Linux 2023** — Com a versão 7.0, os EMR clusters da Amazon agora são executados em AL2 0.23 por padrão. Para obter informações sobre como isso afeta a AMI versão padrão, consulte [Considerações sobre atualização de software](#) no Amazon EMR Management Guide. Observe também que AL2 023 removeu o Python 2.7, portanto, qualquer componente que exija Python agora deve ser escrito com o Python 3.
- **S3 on Outposts com s3a** — A Amazon agora oferece suporte aos EMR buckets Amazon S3 on Outposts com o sistema de arquivos. s3a Para obter mais informações sobre o S3 no Outposts, consulte [O que é o S3 no Outposts?](#) no Guia do usuário do Amazon Simple Storage Service.

Problemas conhecidos

- Observe que você não pode executar mais de uma linha por vez se estiver usando PySpark as versões 3.10 ou posteriores do Python. Você deve executar cada linha, uma de cada vez.

Alterações, melhorias e problemas resolvidos

- Para manter o estado de todas as instâncias em um cluster, a Amazon EMR se integra com ApacheYARN, Apache e HDFS Kerberos. Com a versão 7.0, aprimoramos essas integrações para remover de forma confiável o estado das instâncias que são encerradas devido ao ajuste de escala ou a outras operações. Isso é particularmente importante para clusters de longa duração que têm o ajuste de escala gerenciado habilitado, pois eles escalam com frequência e acumulam milhares de instâncias encerradas ao longo da vida útil.
- Esta versão melhora a configuração do Kerberos para incluir suporte somente para cifras AES baseadas. Kerberos KDC com cifras não AES baseadas não são mais compatíveis com EMR clusters executados nas EMR versões 7.0.0 e superiores da Amazon. AEScifras baseadas oferecem a maior segurança para seus clusters.
- Como parte da migração AWS SDK 2.x, o Amazon EMR 7.0 inclui uma atualização do conector Spark Kinesis para fins de compatibilidade. Essa atualização não está disponível na versão comunitária do Apache Spark. Se você usa o conector Spark Kinesis de uma versão da EMR Amazon inferior à 7.0, você deve migrar os códigos do seu aplicativo para execução SDK em 2.x antes de poder migrar suas cargas de trabalho para o Amazon 7.0. EMR Para obter mais informações, consulte [Migração do conector Spark Kinesis para SDK 2.x para Amazon 7.0 EMR](#).
- Quando você inicia um cluster com a versão de patch mais recente do Amazon EMR 5.36 ou superior, 6.6 ou superior ou 7.0 ou superior, a Amazon EMR usa a versão mais recente do Amazon Linux 2023 ou Amazon Linux 2 para a Amazon padrão. EMR AMI Para obter mais informações, consulte [Usando o Amazon Linux padrão AMI para a Amazon EMR](#).

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2023.5.2 240708.1	6.1.96-102.177.amzn2023	23 de julho de 2024	Leste dos EUA (Norte da Virgínia), Leste dos

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			<p>EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (UAE Central), Canadá (UAE Central), Israel (Tel Aviv), Oeste do Canadá (Calgary), (Oeste dos EUA), AWS GovCloud</p>

OsRelease Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2023.3.2 240304.0	6.1.79-99.164.amzn2023	12 de março de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), Oeste do Canadá (Calgary), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2023.3.2 240219.0	6.1.77-99.164.amzn2023	1.º de março de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (Doha), Canadá (UAE Central),

OsRelease Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), Oeste do Canadá (Calgary), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2023.3.2 240205.0	6.1.75-99.163.amzn2023	19 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), Oeste do Canadá (Calgary), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2023.3.2 240122.0	6.1.72-96.166.amzn2023	5 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (Doha), Canadá (UAE Central),

OsRelease Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), Oeste do Canadá (Calgary), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2023.3.2 240108.0	6.1.72-96.166.amzn2023	24 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (UAE Central), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), Oeste do Canadá (Calgary), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2023.3.2 231211.4	6.1.66-91.160.amzn2023	19 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

Versões padrão do Java 7.0.0

A EMR versão 7.0 e superior da Amazon vem com o Amazon Corretto 17 (construído no JDK Open) por padrão para aplicativos compatíveis com o Corretto 17 JDK (17), com exceção do Apache Livy.

A tabela a seguir mostra as versões padrão do Java para aplicativos na Amazon EMR 7.0.0. Se você quiser alterar o padrão JVM em seu cluster, siga as [Configurar aplicações para usar uma Máquina Virtual Java específica](#) instruções em cada aplicativo executado no cluster. Você só pode usar uma versão de runtime do Java para um cluster. A Amazon EMR não oferece suporte à execução de diferentes nós ou aplicativos em diferentes versões de tempo de execução no mesmo cluster.

Aplicativo	Versão Java/Amazon Corretto (o padrão é negrito)
Delta	17, 11, 8
Flink	11, 8
HBase	11, 8
HCatalog	17, 11, 8
Hadoop	17, 11, 8
Hive	17, 11, 8

Aplicativo	Versão Java/Amazon Corretto (o padrão é negrito)
Hudi	17, 11, 8
Iceberg	17, 11, 8
Livy	17, 11, 8
Oozie	17, 11, 8
Phoenix	8
PrestoDB	8
Spark	17, 11, 8
Faísca RAPIDS	17, 11, 8
Sqoop	8
Tez	17, 11, 8
Trino	17
Zeppelin	8
Pig	8
Zookeeper	8

Versões de componente 7.0.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado myapp-component com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como 2.2-amzn-2.

Componente	Version (Versão)	Descrição
adot-java-agent	1.31.0	Um Java Agent que coleta métricas de daemons de aplicações.
delta	3.0.0	O Delta Lake é um formato de tabela aberto para grandes conjuntos de dados analíticos
delta-standalone-connectors	0.6.0	Os conectores do Delta oferecem runtimes diferentes para integrar o Delta Lake a mecanismos, como Flink, Hive e Presto.
emr-amazon-cloudwatch-agent	1.300031.1-amzn-0	Um aplicativo que coleta métricas internas em nível de sistema e métricas de aplicativos personalizadas de instâncias da Amazon. EC2
emr-ddb	5.2.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	3.9.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.13.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
emr-notebook-env	1.7.0	Ambiente Conda para bloco de anotações do emr, que inclui o jupyter enterprise gateway
emr-s3-dist-cp	2.30.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	2.9.0	EMRConector S3Select
emr-wal-cli	1.2.1	Cli usada para lista/exclusão de emrwal.
emrfs	2.61.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.18.0-amzn-0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
flink-jobmanager-config	1.18.0-amzn-0	Gerenciando recursos em EMR nós para o Apache JobManager Flink.
hadoop-client	3.3.6-amzn-2	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	3.3.6-amzn-2	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	3.3.6-amzn-2	HDFScliente e biblioteca de linha de comando

Componente	Version (Versão)	Descrição
hadoop-hdfs-namenode	3.3.6-amzn-2	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	3.3.6-amzn-2	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httfs-server	3.3.6-amzn-2	HTTPendpoint para HDFS operações.
hadoop-kms-server	3.3.6-amzn-2	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	3.3.6-amzn-2	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	3.3.6-amzn-2	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	3.3.6-amzn-2	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	3.3.6-amzn-2	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	2.4.17-amzn-4	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.

Componente	Version (Versão)	Descrição
hbase-region-server	2.4.17-amzn-4	Serviço para atender uma ou mais HBase regiões.
hbase-client	2.4.17-amzn-4	HBase cliente de linha de comando.
hbase-rest-server	2.4.17-amzn-4	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	2.4.17-amzn-4	Serviço que fornece um endpoint Thrift para HBase
hbase-operator-tools	2.4.17-amzn-4	Ferramenta de reparo para HBase clusters Apache.
hcatalog-client	3.1.3-amzn-9	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	3.1.3-amzn-9	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	3.1.3-amzn-9	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	3.1.3-amzn-9	O cliente da linha de comando do Hive.
hive-hbase	3.1.3-amzn-9	Cliente do Hive-hbase.

Componente	Version (Versão)	Descrição
hive-metastore-server	3.1.3-amzn-9	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	3.1.3-amzn-9	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.14.0-amzn-1	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.
hudi-presto	0.14.0-amzn-1	Biblioteca de pacotes para executar o Presto com o Hudi.
hudi-trino	0.14.0-amzn-1	Biblioteca de pacotes para execução do Presto com o Hudi.
hudi-spark	0.14.0-amzn-1	Biblioteca de pacotes para executar o Spark com o Hudi.
hue-server	4.11.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
iceberg	q1.4.2-amzn-0	Apache Iceberg é um formato de tabela aberta para conjuntos de dados analíticos imensos
jupyterhub	1.5.0	Servidor de vários usuários para blocos de anotações Jupyter

Componente	Version (Versão)	Descrição
livy-server	0.7.1-incubating	REST interface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mxnet	1.9.1	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mariadb-server	5.5.68+	Servidor de banco de dados MariaDB.
nvidia-cuda	11.8.0	Drivers NVIDIA e toolkit CUDA
oozie-client	5.2.1	O cliente da linha de comando do Oozie.
oozie-server	5.2.1	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	4.7.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	5.1.3	As bibliotecas do phoenix para servidor e cliente
phoenix-connectors	5.1.3	Conectores do Apache Phoenix para Spark-3
phoenix-query-server	5.1.3	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API

Componente	Version (Versão)	Descrição
presto-coordinator	0.283-amzn-1	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.283-amzn-1	O serviço que executa partes de uma consulta.
presto-client	0.283-amzn-1	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
trino-coordinator	426-amzn-1	Serviço que aceita consultas e gerencia a execução de consultas entre trino-workers.
trino-worker	426-amzn-1	O serviço que executa partes de uma consulta.
trino-client	426-amzn-1	Cliente de linha de comando do Trino instalado em mestres de espera de um cluster de alta disponibilidade (HA) em que o servidor do Trino não é iniciado.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	4.1.3	O projeto R para computação estatística
ranger-kms-server	2.0.0	Sistema de gerenciamento de chaves Apache Ranger

Componente	Version (Versão)	Descrição
spark-client	3.5.0-amzn-0	Os clientes da linha de comando do Spark.
spark-history-server	3.5.0-amzn-0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	3.5.0-amzn-0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	3.5.0-amzn-0	Bibliotecas do Apache Spark necessárias para escravos. YARN
spark-rapids	23.10.0-amzn-0	RAPIDSPlugin Nvidia Spark que acelera o Apache Spark com o. GPUs
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	2.11.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.10.2-amzn-7	O YARN aplicativo e as bibliotecas tez.
tez-on-worker	0.10.2-amzn-7	O YARN aplicativo tez e as bibliotecas para nós de trabalho.
webserver	2.4.58	HTTPServidor Apache.

Componente	Version (Versão)	Descrição
zeppelin-server	0.10.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.5.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.5.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 7.0.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para obter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-7.0.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
container-executor	Altere os valores no arquivo <code>container-executor.cfg</code> YARN do Hadoop.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.

Classificações	Descrição
docker-conf	Alterar as configurações relacionadas ao docker.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-session	Alterar as configurações de log4j-yarn-session.properties no Flink para sessão do Kubernetes/Yarn.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.

Classificações	Descrição
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-env	Mude os valores no HDFS ambiente.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive	Configurações EMR selecionadas pela Amazon para o Apache Hive.
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.

Classificações	Descrição
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-java-home	Mude a página inicial do Java do KMS Hadoop
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.

Classificações	Descrição
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
hudi-env	Altere os valores no ambiente do Hudi.
hudi-defaults	Alterar os valores no arquivo hudi-defaults.conf do Hudi.
iceberg-defaults	Alterar os valores no arquivo iceberg-defaults.conf do Iceberg.
delta-defaults	Alterar os valores no arquivo delta-defaults.conf do Delta.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j2	Alterar as configurações de log4j2.properties do Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.

Classificações	Descrição
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.

Classificações	Descrição
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-lakeformation	Alterar os valores no arquivo lakeformation.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.

Classificações	Descrição
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.
trino-log	Alterar os valores no arquivo log.properties do Trino.
trino-config	Alterar os valores no arquivo config.properties do Trino.
trino-password-authenticator	Alterar os valores no arquivo password-authenticator.properties do Trino.
trino-env	Alterar os valores no arquivo trino-env.sh do Trino.
trino-node	Alterar os valores no arquivo node.properties do Trino.
trino-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Trino.
trino-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Trino.
trino-connector-delta	Alterar os valores no arquivo delta.properties do Trino.

Classificações	Descrição
trino-connector-hive	Alterar os valores no arquivo hive.properties do Trino.
trino-exchange-manager	Alterar os valores no arquivo exchange-manager.properties do Trino.
trino-connector-iceberg	Alterar os valores no arquivo iceberg.properties do Trino.
trino-connector-hudi	Alterar os valores no arquivo hudi.properties do Trino.
trino-connector-jmx	Alterar os valores no arquivo jmx.properties do Trino.
trino-connector-kafka	Alterar os valores no arquivo kafka.properties do Trino.
trino-connector-localfile	Alterar os valores no arquivo localfile.properties do Trino.
trino-connector-memory	Alterar os valores no arquivo memory.properties do Trino.
trino-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Trino.
trino-connector-mysql	Alterar os valores no arquivo mysql.properties do Trino.
trino-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Trino.
trino-connector-raptor	Alterar os valores no arquivo raptor.properties do Trino.
trino-connector-redis	Alterar os valores no arquivo redis.properties do Trino.

Classificações	Descrição
trino-connector-redshift	Alterar os valores no arquivo redshift.properties do Trino.
trino-connector-tpch	Alterar os valores no arquivo tpch.properties do Trino.
trino-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Trino.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.
ranger-kms-logback	Altere os valores no arquivo kms-logback.xml do Ranger. KMS
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j2	Alterar os valores no arquivo log4j2.properties do Spark.

Classificações	Descrição
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zeppelin-site	Alterar as definições da configuração em zeppelin-site.xml.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Log de alterações 7.0.0

Log de alterações da versão 7.0.0 e das notas da versão

Data	Evento	Descrição
2023-12-29	Implantação concluída	Amazon EMR 7.0.0 totalmente implantado em todas as regiões suportadas
2023-12-28	Publicação de documentos	Notas de lançamento do Amazon EMR 7.0.0 publicadas pela primeira vez
2023-12-19	Lançamento inicial	Amazon EMR 7.0.0 foi implantado pela primeira vez nas regiões comerciais iniciais

Versões de lançamento do Amazon EMR 6.x

Esta seção contém versões de aplicativos, notas de lançamento, versões de componentes e classificações de configuração disponíveis em cada versão de lançamento do Amazon EMR 6.x.

Ao iniciar um cluster, você pode escolher entre várias versões da AmazonEMR. Isso permite que você teste e use versões de aplicações que atendem aos seus requisitos de compatibilidade. Você especifica o número da versão com o rótulo da versão. Os rótulos de versão estão no formato `emr-x.x.x`. Por exemplo, `emr-7.2.0`.

Novos EMR lançamentos da Amazon são disponibilizados em diferentes regiões por um período de vários dias, começando com a primeira região na data de lançamento inicial. A versão mais recente pode não estar disponível em sua região durante esse período.

Para obter uma tabela abrangente das versões do aplicativo em cada versão EMR 6.x da Amazon, consulte [Versões de aplicações nas versões 6.x do Amazon EMR](#).

Tópicos

- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Amazon EMR versão 6.15.0](#)

- [Amazon EMR versão 6.14.0](#)
- [Amazon EMR versão 6.13.0](#)
- [Amazon EMR versão 6.12.0](#)
- [Amazon EMR versão 6.11.1](#)
- [Amazon EMR versão 6.11.0](#)
- [Amazon EMR versão 6.10.1](#)
- [Amazon EMR versão 6.10.0](#)
- [Amazon EMR versão 6.9.1](#)
- [Amazon EMR versão 6.9.0](#)
- [Amazon EMR versão 6.8.1](#)
- [Amazon EMR versão 6.8.0](#)
- [Amazon EMR versão 6.7.0](#)
- [Amazon EMR versão 6.6.0](#)
- [Amazon EMR versão 6.5.0](#)
- [Amazon EMR versão 6.4.0](#)
- [Amazon EMR versão 6.3.1](#)
- [Amazon EMR versão 6.3.0](#)
- [Amazon EMR versão 6.2.1](#)
- [Amazon EMR versão 6.2.0](#)
- [Amazon EMR versão 6.1.1](#)
- [Amazon EMR versão 6.1.0](#)
- [EMR Versão 6.0.1 da Amazon](#)
- [Amazon EMR versão 6.0.0](#)

Versões de aplicações nas versões 6.x do Amazon EMR

Para obter uma tabela abrangente que liste as versões das aplicações disponíveis em cada versão 6.x do Amazon EMR, abra as [versões de aplicações nas versões 6.x do Amazon EMR](#) no seu navegador.

Amazon EMR versão 6.15.0

Versões 6.15.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-6.15.0	emr-6.14.0	emr-6.13.0	emr-6.12.0
AWS SDK para 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	2.12.17	2.12.15	2.12.15	2.12.15
AmazonCloudWatchAgent	-	-	-	-
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

	emr-6.15.0	emr-6.14.0	emr-6.13.0	emr-6.12.0
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
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	0.10.2	0.10.2	0.10.2	0.10.2

	emr-6.15.0	emr-6.14.0	emr-6.13.0	emr-6.12.0
Trino (PrestoSQL)	426	422	414	414
Zeppelin	0.10.1	0.10.1	0.10.1	0.10.1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

Notas da versão 6.15.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 6.15.0 da Amazon. As alterações são referentes à versão 6.14.0. Para obter informações sobre o cronograma da versão, consulte o [Log de alterações 6.15.0](#).

Novos atributos

- Atualizações da aplicação: 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.
- [Lançamentos mais rápidos para EMR clusters executados em EC2](#) — Agora é até 35% mais rápido lançar um Amazon EMR on EC2 cluster. Com essa melhoria, a maioria dos clientes pode lançar seus clusters em cinco minutos ou menos.
- [CodeWhisperer para EMR Studio](#) — Agora você pode usar a Amazon CodeWhisperer com o Amazon EMR Studio para obter recomendações em tempo real à medida que escreve código JupyterLab. CodeWhisperer pode concluir seus comentários, concluir linhas únicas de código, fazer line-by-line recomendações e gerar funções totalmente formadas.
- [Tempos mais rápidos de reinício do trabalho com o Flink](#) — Com o Amazon EMR 6.15.0 e superior, vários novos mecanismos estão disponíveis para o Apache Flink para melhorar o tempo de reinício do trabalho durante as operações de recuperação ou escalabilidade de tarefas. Isso otimiza a velocidade de recuperação e reinicialização dos gráficos de execução para melhorar a estabilidade do trabalho.
- [Controle de acesso detalhado e em nível de tabela para formatos de tabela aberta — Com o EMR Amazon 6.15.0 e superior, quando você executa trabalhos do Spark na Amazon em EC2 clusters que acessam dados EMR no AWS Glue Data Catalog, você pode AWS Lake Formation usar para aplicar permissões em nível de tabela, linha, coluna e célula em tabelas baseadas em Hudi, Iceberg ou Delta Lake.](#)

- **Atualização do Hadoop** — O Amazon EMR 6.15.0 inclui uma atualização do Apache Hadoop para a versão 3.3.6. O Hadoop 3.3.6 era a versão mais recente na época da implantação do Amazon EMR 6.15, lançada pela Apache em junho de 2023. Versões anteriores da Amazon EMR (6.9.0 a 6.14.x) usavam o Hadoop 3.3.3.

A atualização inclui centenas de melhorias e correções, além de recursos que incluem parâmetros de nós de dados reconfiguráveis, DFSAdmin opção de iniciar operações de reconfiguração em massa em todos os datanodes ativos e um vetor que permite API que leitores exigentes especifiquem vários intervalos para leitura. O Hadoop 3.3.6 também adiciona suporte HDFS APIs e semântica ao seu registro de gravação antecipada (WAL), para que ele HBase possa ser executado em outras implementações de sistemas de armazenamento. Para obter mais informações, consulte os logs de alterações das versões [3.3.4](#), [3.3.5](#) e [3.3.6](#) na documentação do Apache Hadoop.

- **Support AWS SDK for Java, versão 2** - Os aplicativos Amazon EMR 6.15.0 podem ser usados AWS SDK para Java nas versões [1.12.569](#) ou [2.20.160](#) se o aplicativo oferecer suporte à v2. O AWS SDK for Java 2.x é uma grande reescrita da base de código da versão 1.x. Ele foi criado com base no Java 8+ e adiciona vários recursos frequentemente solicitados. Isso inclui suporte para E/S sem bloqueio e a capacidade de conectar uma HTTP implementação diferente em tempo de execução. Para obter mais informações, incluindo um Guia de migração de SDK Java v1 para v2, consulte o guia [AWS SDK para Java, versão 2](#).

Alterações, melhorias e problemas resolvidos

- Para melhorar seus EMR clusters de alta disponibilidade, esta versão permite a conectividade com EMR daemons da Amazon em um host local que usa endpoints. IPv6
- Esta versão permite a comunicação TLS 1.2 com ZooKeeper provisionados em todos os nós primários do seu cluster de alta disponibilidade.
- Esta versão melhora o gerenciamento dos arquivos de log de ZooKeeper transações que são mantidos nos nós primários para minimizar os cenários em que os arquivos de log ultrapassam os limites e interrompem as operações do cluster.
- Essa versão torna a comunicação entre nós mais resiliente para clusters de alta disponibilidade. EMR Essa melhoria reduz a chance de falhas na ação de bootstrap ou na inicialização do cluster.
- O Tez no Amazon EMR 6.15.0 introduz configurações que você pode especificar para abrir de forma assíncrona as divisões de entrada em uma divisão agrupada Tez. Isso resulta em uma performance mais rápida das consultas de leitura quando há um grande número de divisões de

entrada em uma única divisão agrupada do Tez. Para obter mais informações, consulte a [Abertura assíncrona da divisão do Tez](#).

- Quando você inicia um cluster com a versão de patch mais recente do Amazon EMR 5.36 ou superior, 6.6 ou superior ou 7.0 ou superior, a Amazon EMR usa a versão mais recente do Amazon Linux 2023 ou Amazon Linux 2 para a Amazon padrão. EMR AMI Para obter mais informações, consulte [Usando o Amazon Linux padrão AMI para a Amazon EMR](#).

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 709.1	4.14.348	23 de julho de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Médio (Bahrein), Canadá (Central), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia), Ásia Pacífico (Hyderabad), Oriente Médio (), Europa (EspanhaUAE), Europa (Zurique), Ásia-Pacífico (Melbourne), Israel (Tel Aviv), Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 223.0	4.14.336	8 de março de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia), Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 131.0	4.14.336	14 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia) , Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 124.0	4.14.336	7 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia), Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 109.0	4.14.334	24 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia) , Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 218.0	4.14.330	2 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 206.0	4.14.330	22 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 116.0	4.14.328	11 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 101.0	4.14.327	13 de novembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

Versões padrão do Java 6.15.0

As EMR versões 6.12.0 e superiores da Amazon oferecem suporte a todos os aplicativos com o Amazon Corretto 8 por padrão, exceto o Trino. Para Trino, a Amazon EMR oferece suporte ao Amazon Corretto 17 por padrão, começando com a versão 6.9.0 da EMR Amazon. A Amazon EMR também oferece suporte a alguns aplicativos com o Amazon Corretto 11 e 17. Essas aplicações estão listadas na tabela a seguir. Se você quiser alterar o padrão JVM em seu cluster, siga as [Configurar aplicações para usar uma Máquina Virtual Java específica](#) instruções em cada aplicativo executado no cluster. Você só pode usar uma versão de runtime do Java para um cluster. A Amazon EMR não oferece suporte à execução de diferentes nós ou aplicativos em diferentes versões de tempo de execução no mesmo cluster.

Embora a Amazon EMR ofereça suporte ao Amazon Corretto 11 e 17 no Apache Spark, no Apache Hadoop e no Apache Hive, o desempenho pode regredir em algumas cargas de trabalho quando você usa essas versões do Corretto. Recomendamos que você teste suas workloads antes de alterar os padrões.

A tabela a seguir mostra as versões padrão do Java para aplicativos na Amazon EMR 6.15.0:

Aplicativo	Versão Java/Amazon Corretto (o padrão é negrito)
Delta	17, 11, 8
Flink	11, 8

Aplicativo	Versão Java/Amazon Corretto (o padrão é negrito)
Ganglia	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
Faísca RAPIDS	17, 11, 8
Sqoop	8
Tez	17, 11, 8
Trino	17
Zeppelin	8
Pig	8
Zookeeper	8

Versões do componente 6.15.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.4.2	Amazon SageMaker Spark SDK
<code>delta</code>	2.4.0	O Delta Lake é um formato de tabela aberto para grandes conjuntos de dados analíticos
<code>delta-standalone-connectors</code>	0.6.0	Os conectores do Delta oferecem runtimes diferentes para integrar o Delta Lake a mecanismos, como Flink, Hive e Presto.
<code>emr-ddb</code>	5.2.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	3.8.0	Bibliotecas convenientes para o ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
emr-kinesis	3.12.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-notebook-env	1.7.0	Ambiente Conda para bloco de anotações do emr, que inclui o jupyter enterprise gateway
emr-s3-dist-cp	2.29.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	2.8.0	EMRConector S3Select
emr-wal-cli	1.2.0	Cli usada para lista/exclusão de emrwal.
emrfs	2.60,0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.17.1-amzn-1	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
flink-jobmanager-config	1.17.1-amzn-1	Gerenciando recursos em EMR nós para o Apache JobManager Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.

Componente	Version (Versão)	Descrição
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	3.3.6-amzn-1	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	3.3.6-amzn-1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	3.3.6-amzn-1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	3.3.6-amzn-1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	3.3.6-amzn-1	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httpfs-server	3.3.6-amzn-1	HTTPendpoint para HDFS operações.
hadoop-kms-server	3.3.6-amzn-1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API

Componente	Version (Versão)	Descrição
hadoop-mapred	3.3.6-amzn-1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	3.3.6-amzn-1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	3.3.6-amzn-1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	3.3.6-amzn-1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	2.4.17-amzn-3	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	2.4.17-amzn-3	Serviço para atender a uma ou mais HBase regiões.
hbase-client	2.4.17-amzn-3	HBasecliente de linha de comando.
hbase-rest-server	2.4.17-amzn-3	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	2.4.17-amzn-3	Serviço que fornece um endpoint Thrift para. HBase

Componente	Version (Versão)	Descrição
hbase-operator-tools	2.4.17-amzn-3	Ferramenta de reparo para HBase clusters Apache.
hcatalog-client	3.1.3-amzn-8	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	3.1.3-amzn-8	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	3.1.3-amzn-8	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	3.1.3-amzn-8	O cliente da linha de comando do Hive.
hive-hbase	3.1.3-amzn-8	Cliente do Hive-hbase.
hive-metastore-server	3.1.3-amzn-8	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	3.1.3-amzn-8	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.14.0-amzn-0	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.

Componente	Version (Versão)	Descrição
hudi-presto	0.14.0-amzn-0	Biblioteca de pacotes para executar o Presto com o Hudi.
hudi-trino	0.14.0-amzn-0	Biblioteca de pacotes para execução do Presto com o Hudi.
hudi-spark	0.14.0-amzn-0	Biblioteca de pacotes para executar o Spark com o Hudi.
hue-server	4.11.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
iceberg	1.4.0-amzn-0	Apache Iceberg é um formato de tabela aberta para conjuntos de dados analíticos imensos
jupyterhub	1.5.0	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.7.1-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mxnet	1.9.1	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mariadb-server	5.5.68+	Servidor de banco de dados MariaDB.
nvidia-cuda	11.8.0	Drivers NVIDIA e toolkit CUDA

Componente	Version (Versão)	Descrição
oozie-client	5.2.1	O cliente da linha de comando do Oozie.
oozie-server	5.2.1	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	4.7.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	5.1.3	As bibliotecas do phoenix para servidor e cliente
phoenix-connectors	5.1.3	Conectores do Apache Phoenix para Spark-3
phoenix-query-server	5.1.3	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.283-amzn-0	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.283-amzn-0	O serviço que executa partes de uma consulta.
presto-client	0.283-amzn-0	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.

Componente	Version (Versão)	Descrição
trino-coordinator	426-amzn-0	Serviço que aceita consultas e gerencia a execução de consultas entre trino-workers.
trino-worker	426-amzn-0	O serviço que executa partes de uma consulta.
trino-client	426-amzn-0	Cliente de linha de comando do Trino instalado em mestres de espera de um cluster de alta disponibilidade (HA) em que o servidor do Trino não é iniciado.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	4.0.2	O projeto R para computação estatística
ranger-kms-server	2.0.0	Sistema de gerenciamento de chaves Apache Ranger
spark-client	3.4.1-amzn-2	Os clientes da linha de comando do Spark.
spark-history-server	3.4.1-amzn-2	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	3.4.1-amzn-2	Mecanismo de execução na memória para YARN.

Componente	Version (Versão)	Descrição
spark-yarn-slave	3.4.1-amzn-2	Bibliotecas do Apache Spark necessárias para escravos. YARN
spark-rapids	23.08.1-amzn-0	RAPIDSPlugin Nvidia Spark que acelera o Apache Spark com o. GPUs
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	2.11.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.10.2-amzn-6	O YARN aplicativo e as bibliotecas tez.
tez-on-worker	0.10.2-amzn-6	O YARN aplicativo tez e as bibliotecas para nós de trabalho.
webserver	2.4.41+	HTTPServidor Apache.
zeppelin-server	0.10.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.5.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.

Componente	Version (Versão)	Descrição
zookeeper-client	3.5.10	ZooKeeper cliente de linha de comando.

Classificações da configuração 6.15.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Ocorrem ações de reconfiguração quando você especifica uma configuração para grupos de instâncias em um cluster em execução. A Amazon EMR só inicia ações de reconfiguração para as classificações que você modifica. Para ter mais informações, consulte [Reconfigurar um grupo de instâncias de um cluster em execução](#).

Classificações do emr-6.15.0

Classificações	Descrição	Ações de reconfiguração
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.	Restarts the Resource Manager service.
container-executor	Altere os valores no arquivo <code>container-executor.cfg</code> YARN do Hadoop.	Not available.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.	Not available.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager,

Classificações	Descrição	Ações de reconfiguração
		NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Alterar as configurações relacionadas ao docker.	Not available.
emrfs-site	Altere EMRFS as configurações.	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	Alterar as configurações do flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Alterar as configurações de log4j.properties no Flink.	Restarts Flink history server.

Classificações	Descrição	Ações de reconfiguração
flink-log4j-session	Alterar as configurações de log4j-yarn-session.properties no Flink para sessão do Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.	Restarts Flink history server.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do 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	Alterar os valores no arquivo log4j.properties do 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.

Classificações	Descrição	Ações de reconfiguração
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop	Not available.
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop	Not available.
hbase	Configurações EMR seleciona das pela Amazon para Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Mude os valores no HBase ambiente.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.	Not available.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.	This classification should not be reconfigured.

Classificações	Descrição	Ações de reconfiguração
hdfs-env	Mude os valores no HDFS ambiente.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Mude os valores no HCatalog ambiente.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Configurações EMR selecionadas pela Amazon para o Apache Hive.	Sets configurations to launch Hive LLAP service.

Classificações	Descrição	Ações de reconfiguração
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.	Not available.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.	Not available.
hive-env	Alterar os valores no ambiente do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log 4j2.properties do Hive.	Not available.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log 4j2.properties do Hive.	Not available.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.	Not available.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.	Not available.

Classificações	Descrição	Ações de reconfiguração
hue-ini	Alterar os valores no arquivo ini do Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Mude os valores no HTTPFS ambiente.	Restarts Hadoop Httpfs service.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.	Not available.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.	Restarts Hadoop-KMS service.
hadoop-kms-java-home	Mude a página inicial do Java do KMS Hadoop	Not available.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.	Not available.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Altere os valores no ambiente do Hudi.	Not available.
hudi-defaults	Alterar os valores no arquivo hudi-defaults.conf do Hudi.	Not available.
iceberg-defaults	Alterar os valores no arquivo iceberg-defaults.conf do Iceberg.	Not available.

Classificações	Descrição	Ações de reconfiguração
delta-defaults	Alterar os valores no arquivo delta-defaults.conf do Delta.	Not available.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.	Not available.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.	Not available.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.	Not available.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.	Not available.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.	Restarts Livy Server.
livy-env	Alterar os valores no ambiente do Livy.	Restarts Livy Server.
livy-log4j2	Alterar as configurações de log4j2.properties do Livy.	Restarts Livy Server.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Alterar os valores no ambiente do Oozie.	Restarts Oozie.

Classificações	Descrição	Ações de reconfiguração
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.	Restarts Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.	Not available.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.	Not available.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.	Not available.
pig-env	Alterar os valores no ambiente do Pig.	Not available.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.	Restarts Oozie.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.	Not available.
presto-log	Alterar os valores no arquivo log.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Alterar os valores no arquivo config.properties do Presto.	Restarts Presto-Server (for PrestoDB)

Classificações	Descrição	Ações de reconfiguração
presto-password-authenticator	Altere os valores no arquivo <code>password-authenticator.properties</code> do Presto.	Not available.
presto-env	Altere os valores no arquivo <code>presto-env.sh</code> do Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Altere os valores no arquivo <code>node.properties</code> do Presto.	Not available.
presto-connector-blackhole	Alterar os valores no arquivo <code>blackhole.properties</code> do Presto.	Not available.
presto-connector-cassandra	Alterar os valores no arquivo <code>cassandra.properties</code> do Presto.	Not available.
presto-connector-hive	Alterar os valores no arquivo <code>hive.properties</code> do Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Alterar os valores no arquivo <code>jmx.properties</code> do Presto.	Not available.
presto-connector-kafka	Alterar os valores no arquivo <code>kafka.properties</code> do Presto.	Not available.
presto-connector-lakeformation	Alterar os valores no arquivo <code>lakeformation.properties</code> do Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Alterar os valores no arquivo <code>localfile.properties</code> do Presto.	Not available.
presto-connector-memory	Alterar os valores no arquivo <code>memory.properties</code> do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.	Not available.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.	Not available.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.	Not available.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.	Not available.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.	Not available.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.	Not available.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.	Not available.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.	Not available.
trino-log	Alterar os valores no arquivo log.properties do Trino.	Restarts Trino-Server (for Trino)
trino-config	Alterar os valores no arquivo config.properties do Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Alterar os valores no arquivo password-authenticator.properties do Trino.	Restarts Trino-Server (for Trino)

Classificações	Descrição	Ações de reconfiguração
trino-env	Alterar os valores no arquivo trino-env.sh do Trino.	Restarts Trino-Server (for Trino)
trino-node	Alterar os valores no arquivo node.properties do Trino.	Not available.
trino-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Trino.	Not available.
trino-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Trino.	Not available.
trino-connector-delta	Alterar os valores no arquivo delta.properties do Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Alterar os valores no arquivo hive.properties do Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Alterar os valores no arquivo exchange-manager.properties do Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Alterar os valores no arquivo iceberg.properties do Trino.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Alterar os valores no arquivo hudi.properties do Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Alterar os valores no arquivo jmx.properties do Trino.	Not available.
trino-connector-kafka	Alterar os valores no arquivo kafka.properties do Trino.	Not available.
trino-connector-localfile	Alterar os valores no arquivo localfile.properties do Trino.	Not available.

Classificações	Descrição	Ações de reconfiguração
trino-connector-memory	Alterar os valores no arquivo <code>memory.properties</code> do Trino.	Not available.
trino-connector-mongodb	Alterar os valores no arquivo <code>mongodb.properties</code> do Trino.	Not available.
trino-connector-mysql	Alterar os valores no arquivo <code>mysql.properties</code> do Trino.	Not available.
trino-connector-postgresql	Alterar os valores no arquivo <code>postgresql.properties</code> do Trino.	Not available.
trino-connector-raptor	Alterar os valores no arquivo <code>raptor.properties</code> do Trino.	Not available.
trino-connector-redis	Alterar os valores no arquivo <code>redis.properties</code> do Trino.	Not available.
trino-connector-redshift	Alterar os valores no arquivo <code>redshift.properties</code> do Trino.	Not available.
trino-connector-tpch	Alterar os valores no arquivo <code>tpch.properties</code> do Trino.	Not available.
trino-connector-tpcds	Alterar os valores no arquivo <code>tpcds.properties</code> do Trino.	Not available.
ranger-kms-dbks-site	Altere os valores no arquivo <code>dbks-site.xml</code> do Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-site	Altere os valores no <code>ranger-kms-site</code> arquivo.xml do RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.	Restarts Ranger KMS Server.

Classificações	Descrição	Ações de reconfiguração
ranger-kms-logback	Altere os valores no arquivo kms-logback.xml do Ranger. KMS	Not available.
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.	Not available.
spark	Configurações EMR seleciona das pela Amazon para o Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Alterar os valores no ambiente do Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.	Not available.
spark-log4j2	Alterar os valores no arquivo log4j2.properties do Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Alterar os valores no ambiente do Sqoop.	Not available.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.	Not available.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.	Not available.

Classificações	Descrição	Ações de reconfiguração
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.	Restart Oozie and HiveServer2.
yarn-env	Mude os valores no YARN ambiente.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.	Restarts Zeppelin.
zeppelin-site	Alterar as definições da configuração em zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.	Restarts Zookeeper server.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.	Restarts Zookeeper server.

Log de alterações 6.15.0

Log de alterações da versão 6.15.0 e das notas da versão

Data	Evento	Descrição
2023-11-17	Publicação de documentos	Notas de lançamento do Amazon EMR 6.15.0 publicadas pela primeira vez
2023-11-17	Implantação concluída	Amazon EMR 6.15.0 totalmente implantado em todas as regiões suportadas
2023-11-13	Lançamento inicial	Amazon EMR 6.15.0 foi implantado pela primeira vez nas regiões comerciais iniciais

Amazon EMR versão 6.14.0

Versões da aplicação 6.14.0

As seguintes aplicações são compatíveis com esta versão: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-6.14.0	emr-6.13.0	emr-6.12.0	emr-6.11.1
AWS SDK para 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	2.12.15	2.12.15	2.12.15	2.12.15
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	0.10.2	0.10.2	0.10.2	0.10.2
Trino (PrestoSQL)	422	414	414	410
Zeppelin	0.10.1	0.10.1	0.10.1	0.10.1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

Notas da versão 6.14.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 6.14.0 da Amazon. As alterações são referentes à versão 6.13.0. Para obter informações sobre o cronograma da versão, consulte o [Log de alterações 6.14.0](#).

Novos atributos

- 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.
- A [escalabilidade EMR gerenciada da Amazon](#) agora está disponível na região ap-southeast-3 Ásia-Pacífico (Jacarta) para clusters que você cria com o Amazon EMR 6.14.0 ou superior.

Alterações, melhorias e problemas resolvidos

- A versão 6.14.0 otimiza o gerenciamento de registros com a Amazon em execução EMR na Amazon. EC2 Como resultado, é possível ver uma pequena redução nos custos de armazenamento dos logs do cluster.
- A versão 6.14.0 melhora o fluxo de trabalho de escalabilidade para considerar diferentes instâncias principais que têm uma variação substancial no tamanho de seus volumes da Amazon. EBS Essa melhoria se aplica somente aos nós centrais; as operações de redução dos nós de tarefas não são afetadas.
- A versão 6.14.0 melhora a forma como a Amazon EMR interage com aplicativos de código aberto, como. Apache Hadoop YARN ResourceManager and HDFS NameNode Essa melhoria reduz o risco de atrasos operacionais com o escalonamento do cluster e atenua as falhas de inicialização que ocorrem devido a problemas de conectividade com os aplicações de código aberto.
- A versão 6.14.0 otimiza a instalação da aplicação na inicialização do cluster. Isso melhora os tempos de inicialização do cluster para determinadas combinações de EMR aplicativos da Amazon.
- A versão 6.14.0 corrige um problema em que as operações de redução de escala do cluster podem parar quando um cluster executado em um domínio personalizado é reiniciado VPC no núcleo ou no nó da tarefa.
- Quando você inicia um cluster com a versão de patch mais recente do Amazon EMR 5.36 ou superior, 6.6 ou superior ou 7.0 ou superior, a Amazon EMR usa a versão mais recente do Amazon Linux 2023 ou Amazon Linux 2 para a Amazon padrão. EMR AMI Para obter mais informações, consulte [Usando o Amazon Linux padrão AMI para a Amazon EMR](#).

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 709.1	4.14.348	23 de julho de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			<p>(Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia), Ásia Pacífico (Hyderabad), Oriente Médio (), Europa (EspanhaUAE), Europa (Zurique), Ásia-Pacífico (Melbourne), Israel (Tel Aviv), Oeste do Canadá (Calgary)</p>

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 223.0	4.14.336	8 de março de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia) , Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 131.0	4.14.336	14 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia) , Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 124.0	4.14.336	7 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (UAE Central), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia) , Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 109.0	4.14.334	24 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelease Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia) , Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 218.0	4.14.330	2 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 206.0	4.14.330	22 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 116.0	4.14.328	11 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 101.0	4.14.327	17 de novembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelease Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 906.0	4.14.322	11 de setembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)

Versões padrão do Java 6.14.0

As EMR versões 6.12.0 e superiores da Amazon oferecem suporte a todos os aplicativos com o Amazon Corretto 8 por padrão, exceto o Trino. Para Trino, a Amazon EMR oferece suporte ao Amazon Corretto 17 por padrão, começando com a versão 6.9.0 da EMR Amazon. A Amazon EMR também oferece suporte a alguns aplicativos com o Amazon Corretto 11 e 17. Essas aplicações estão listadas na tabela a seguir. Se você quiser alterar o padrão JVM em seu cluster, siga as [Configurar aplicações para usar uma Máquina Virtual Java específica](#) instruções em cada aplicativo executado no cluster. Você só pode usar uma versão de runtime do Java para um cluster. A Amazon EMR não oferece suporte à execução de diferentes nós ou aplicativos em diferentes versões de tempo de execução no mesmo cluster.

Embora a Amazon EMR ofereça suporte ao Amazon Corretto 11 e 17 no Apache Spark, no Apache Hadoop e no Apache Hive, o desempenho pode regredir em algumas cargas de trabalho quando você usa essas versões do Corretto. Recomendamos que você teste suas workloads antes de alterar os padrões.

A tabela a seguir mostra as versões padrão do Java para aplicativos na Amazon EMR 6.14.0:

Aplicativo	Versão Java/Amazon Corretto (o padrão é negrito)
Delta	17, 11, 8
Flink	11, 8
Ganglia	8
HBase	11, 8

Aplicativo	Versão Java/Amazon Corretto (o padrão é negrito)
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
Faísca RAPIDS	17, 11, 8
Sqoop	8
Tez	17, 11, 8
Trino	17
Zeppelin	8
Pig	8
Zookeeper	8

Versões do componente 6.14.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR

e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.4.2	Amazon SageMaker Spark SDK
<code>delta</code>	2.4.0	O Delta Lake é um formato de tabela aberto para grandes conjuntos de dados analíticos
<code>delta-standalone-connectors</code>	0.6.0	Os conectores do Delta oferecem runtimes diferentes para integrar o Delta Lake a mecanismos, como Flink, Hive e Presto.
<code>emr-ddb</code>	5.1.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	3.7.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.11.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-notebook-env</code>	1.7.0	Ambiente Conda para bloco de anotações do emr, que

Componente	Version (Versão)	Descrição
		inclui o jupyter enterprise gateway
emr-s3-dist-cp	2.28.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	2.7.0	EMRConector S3Select
emr-wal-cli	1.1.0	Cli usada para lista/exclusão de emrwal.
emrfs	2.59.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.17.1-amzn-0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
flink-jobmanager-config	1.17.1-amzn-0	Gerenciando recursos em EMR nós para o Apache JobManager Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.

Componente	Version (Versão)	Descrição
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	3.3.3-amzn-6	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	3.3.3-amzn-6	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	3.3.3-amzn-6	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	3.3.3-amzn-6	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	3.3.3-amzn-6	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httpfs-server	3.3.3-amzn-6	HTTPendpoint para HDFS operações.
hadoop-kms-server	3.3.3-amzn-6	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	3.3.3-amzn-6	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.

Componente	Version (Versão)	Descrição
hadoop-yarn-nodemanager	3.3.3-amzn-6	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	3.3.3-amzn-6	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	3.3.3-amzn-6	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	2.4.17-amzn-2	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	2.4.17-amzn-2	Serviço para atender a uma ou mais HBase regiões.
hbase-client	2.4.17-amzn-2	HBasecliente de linha de comando.
hbase-rest-server	2.4.17-amzn-2	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	2.4.17-amzn-2	Serviço que fornece um endpoint Thrift para. HBase
hbase-operator-tools	2.4.17-amzn-2	Ferramenta de reparo para HBase clusters Apache.
hcatalog-client	3.1.3-amzn-7	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.

Componente	Version (Versão)	Descrição
hcatalog-server	3.1.3-amzn-7	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	3.1.3-amzn-7	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	3.1.3-amzn-7	O cliente da linha de comando do Hive.
hive-hbase	3.1.3-amzn-7	Cliente do Hive-hbase.
hive-metastore-server	3.1.3-amzn-7	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	3.1.3-amzn-7	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.13.1-amzn-2	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.
hudi-presto	0.13.1-amzn-2	Biblioteca de pacotes para executar o Presto com o Hudi.
hudi-trino	0.13.1-amzn-2	Biblioteca de pacotes para execução do Presto com o Hudi.

Componente	Version (Versão)	Descrição
hudi-spark	0.13.1-amzn-2	Biblioteca de pacotes para executar o Spark com o Hudi.
hue-server	4.11.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
iceberg	1.3.1-amzn-0	Apache Iceberg é um formato de tabela aberta para conjuntos de dados analíticos imensos
jupyterhub	1.5.0	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.7.1-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mxnet	1.9.1	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mariadb-server	5.5.68+	Servidor de banco de dados MariaDB.
nvidia-cuda	11.8.0	Drivers NVIDIA e toolkit CUDA
oozie-client	5.2.1	O cliente da linha de comando do Oozie.
oozie-server	5.2.1	O serviço que aceita solicitações de fluxo de trabalho do Oozie.

Componente	Version (Versão)	Descrição
opencv	4.7.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	5.1.3	As bibliotecas do phoenix para servidor e cliente
phoenix-connectors	5.1.3	Conectores do Apache Phoenix para Spark-3
phoenix-query-server	5.1.3	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.281-amzn-2	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.281-amzn-2	O serviço que executa partes de uma consulta.
presto-client	0.281-amzn-2	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
trino-coordinator	422-amzn-0	Serviço que aceita consultas e gerencia a execução de consultas entre trino-workers.
trino-worker	422-amzn-0	O serviço que executa partes de uma consulta.

Componente	Version (Versão)	Descrição
trino-client	422-amzn-0	Cliente de linha de comando do Trino instalado em mestres de espera de um cluster de alta disponibilidade (HA) em que o servidor do Trino não é iniciado.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	4.0.2	O projeto R para computação estatística
ranger-kms-server	2.0.0	Sistema de gerenciamento de chaves Apache Ranger
spark-client	3.4.1-amzn-1	Os clientes da linha de comando do Spark.
spark-history-server	3.4.1-amzn-1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	3.4.1-amzn-1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	3.4.1-amzn-1	Bibliotecas do Apache Spark necessárias para escravos. YARN
spark-rapids	23.06.0-amzn-2	RAPIDSPlugin Nvidia Spark que acelera o Apache Spark com o. GPUs

Componente	Version (Versão)	Descrição
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	2.11.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.10.2-amzn-5	O YARN aplicativo e as bibliotecas tez.
tez-on-worker	0.10.2-amzn-5	O YARN aplicativo tez e as bibliotecas para nós de trabalho.
webserver	2.4.41+	HTTPServidor Apache.
zeppelin-server	0.10.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.5.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.5.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 6.14.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Ocorrem ações de reconfiguração quando você especifica uma configuração para grupos de instâncias em um cluster em execução. A Amazon EMR só inicia ações de reconfiguração para as classificações que você modifica. Para ter mais informações, consulte [Reconfigurar um grupo de instâncias de um cluster em execução](#).

Classificações do emr-6.14.0

Classificações	Descrição	Ações de reconfiguração
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.	Restarts the ResourceManager service.
container-executor	Altere os valores no arquivo container-executor.cfg YARN do Hadoop.	Not available.
container-log4j	Altere os valores no arquivo container-log4j.properties YARN do Hadoop.	Not available.
core-site	Alterar os valores no arquivo core-site.xml do 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	Alterar as configurações relacionadas ao docker.	Not available.

Classificações	Descrição	Ações de reconfiguração
emrfs-site	Altere EMRFS as configurações.	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	Alterar as configurações do flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Alterar as configurações de log4j.properties no Flink.	Restarts Flink history server.
flink-log4j-session	Alterar as configurações de log4j-yarn-session.properties no Flink para sessão do Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.	Restarts Flink history server.

Classificações	Descrição	Ações de reconfiguração
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do 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	Alterar os valores no arquivo log4j.properties do 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	Alterar a configuração do servidor SSL no Hadoop	Not available.
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop	Not available.

Classificações	Descrição	Ações de reconfiguração
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Mude os valores no HBase ambiente.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.	Not available.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.	This classification should not be reconfigured.
hdfs-env	Mude os valores no HDFS ambiente.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classificações	Descrição	Ações de reconfiguração
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Mude os valores no HCatalog ambiente.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Configurações EMR selecionadas pela Amazon para o Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.	Not available.

Classificações	Descrição	Ações de reconfiguração
hive-parquet-logging	Alterar os valores no arquivo <code>parquet-logging.properties</code> do Hive.	Not available.
hive-env	Alterar os valores no ambiente do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive <code>schemaTool CLI</code> commands to verify <code>hive-metastore</code> .
hive-exec-log4j2	Altere os valores no arquivo <code>hive-exec-log 4j2.properties</code> do Hive.	Not available.
hive-llap-daemon-log4j2	Altere os valores no arquivo <code>llap-daemon-log 4j2.properties</code> do Hive.	Not available.
hive-log4j2	Alterar os valores no arquivo <code>hive-log4j2.properties</code> do Hive.	Not available.
hive-site	Alterar os valores no arquivo <code>hive-site.xml</code> do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive <code>schemaTool CLI</code> commands to verify <code>hive-metastore</code> . Also restarts Oozie and Zeppelin.
hiveserver2-site	Alterar os valores no arquivo <code>hiveserver2-site.xml</code> do Hive Server2.	Not available.
hue-ini	Alterar os valores no arquivo <code>ini</code> do Hue	Restarts Hue. Also activates Hue <code>config override CLI</code> commands to pick up new configurations.

Classificações	Descrição	Ações de reconfiguração
httpfs-env	Mude os valores no HTTPFS ambiente.	Restarts Hadoop Httpfs service.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.	Not available.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.	Restarts Hadoop-KMS service.
hadoop-kms-java-home	Mude a página inicial do Java do KMS Hadoop	Not available.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.	Not available.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Altere os valores no ambiente do Hudi.	Not available.
hudi-defaults	Alterar os valores no arquivo hudi-defaults.conf do Hudi.	Not available.
iceberg-defaults	Alterar os valores no arquivo iceberg-defaults.conf do Iceberg.	Not available.
delta-defaults	Alterar os valores no arquivo delta-defaults.conf do Delta.	Not available.

Classificações	Descrição	Ações de reconfiguração
jupyter-notebook-conf	Alterar os valores no arquivo <code>jupyter_notebook_config.py</code> do Notebook Jupyter.	Not available.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo <code>jupyterhub_config.py</code> do.	Not available.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.	Not available.
jupyter-sparkmagic-conf	Altere os valores no arquivo <code>config.json</code> do Sparkmagic.	Not available.
livy-conf	Alterar os valores no arquivo <code>livy.conf</code> do Livy.	Restarts Livy Server.
livy-env	Alterar os valores no ambiente do Livy.	Restarts Livy Server.
livy-log4j2	Alterar as configurações de <code>log4j2.properties</code> do Livy.	Restarts Livy Server.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Altere os valores no arquivo <code>mapred-site.xml</code> do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Alterar os valores no ambiente do Oozie.	Restarts Oozie.
oozie-log4j	Alterar os valores no arquivo <code>oozie-log4j.properties</code> do Oozie.	Restarts Oozie.

Classificações	Descrição	Ações de reconfiguração
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.	Not available.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.	Not available.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.	Not available.
pig-env	Alterar os valores no ambiente do Pig.	Not available.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.	Restarts Oozie.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.	Not available.
presto-log	Alterar os valores no arquivo log.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Alterar os valores no arquivo config.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-env	Altere os valores no arquivo presto-env.sh do Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Altere os valores no arquivo node.properties do Presto.	Not available.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.	Not available.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.	Not available.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.	Not available.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.	Not available.
presto-connector-lakeformat ion	Alterar os valores no arquivo lakeformation.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.	Not available.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.	Not available.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.	Not available.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.	Not available.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.	Not available.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.	Not available.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.	Not available.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.	Not available.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.	Not available.
trino-log	Alterar os valores no arquivo log.properties do Trino.	Restarts Trino-Server (for Trino)
trino-config	Alterar os valores no arquivo config.properties do Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Alterar os valores no arquivo password-authenticator.properties do Trino.	Restarts Trino-Server (for Trino)
trino-env	Alterar os valores no arquivo trino-env.sh do Trino.	Restarts Trino-Server (for Trino)

Classificações	Descrição	Ações de reconfiguração
trino-node	Alterar os valores no arquivo <code>node.properties</code> do Trino.	Not available.
trino-connector-blackhole	Alterar os valores no arquivo <code>blackhole.properties</code> do Trino.	Not available.
trino-connector-cassandra	Alterar os valores no arquivo <code>cassandra.properties</code> do Trino.	Not available.
trino-connector-delta	Alterar os valores no arquivo <code>delta.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Alterar os valores no arquivo <code>hive.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Alterar os valores no arquivo <code>exchange-manager.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Alterar os valores no arquivo <code>iceberg.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Alterar os valores no arquivo <code>hudi.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Alterar os valores no arquivo <code>jmx.properties</code> do Trino.	Not available.
trino-connector-kafka	Alterar os valores no arquivo <code>kafka.properties</code> do Trino.	Not available.
trino-connector-localfile	Alterar os valores no arquivo <code>localfile.properties</code> do Trino.	Not available.
trino-connector-memory	Alterar os valores no arquivo <code>memory.properties</code> do Trino.	Not available.

Classificações	Descrição	Ações de reconfiguração
trino-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Trino.	Not available.
trino-connector-mysql	Alterar os valores no arquivo mysql.properties do Trino.	Not available.
trino-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Trino.	Not available.
trino-connector-raptor	Alterar os valores no arquivo raptor.properties do Trino.	Not available.
trino-connector-redis	Alterar os valores no arquivo redis.properties do Trino.	Not available.
trino-connector-redshift	Alterar os valores no arquivo redshift.properties do Trino.	Not available.
trino-connector-tpch	Alterar os valores no arquivo tpch.properties do Trino.	Not available.
trino-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Trino.	Not available.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.	Restarts Ranger KMS Server.
ranger-kms-logback	Altere os valores no arquivo kms-logback.xml do Ranger. KMS	Not available.

Classificações	Descrição	Ações de reconfiguração
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.	Not available.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Alterar os valores no ambiente do Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.	Not available.
spark-log4j2	Alterar os valores no arquivo log4j2.properties do Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Alterar os valores no ambiente do Sqoop.	Not available.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.	Not available.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.	Not available.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.	Restart Oozie and HiveServer2.

Classificações	Descrição	Ações de reconfiguração
yarn-env	Mude os valores no YARN ambiente.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.	Restarts Zeppelin.
zeppelin-site	Alterar as definições da configuração em zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.	Restarts Zookeeper server.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.	Restarts Zookeeper server.

Log de alterações 6.14.0

Log de alterações da versão 6.14.0 e das notas da versão

Data	Evento	Descrição
2023-11-02	Implantação concluída	Amazon EMR 6.14.0 totalmente implantado em todas as regiões suportadas
2023-10-10	Publicação de documentos	Notas de lançamento do Amazon EMR 6.14.0 publicadas pela primeira vez
2023-10-04	Lançamento inicial	Amazon EMR 6.14.0 foi implantado pela primeira vez nas regiões comerciais iniciais

Amazon EMR versão 6.13.0

Versões da aplicação 6.13.0

As seguintes aplicações são compatíveis com esta versão: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-6.13.0	emr-6.12.0	emr-6.11.1	emr-6.11.0
AWS SDK para 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	2.12.15	2.12.15	2.12.15	2.12.15
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
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

	emr-6.13.0	emr-6.12.0	emr-6.11.1	emr-6.11.0
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	0.10.2	0.10.2	0.10.2	0.10.2
Trino (PrestoSQL)	414	414	410	410
Zeppelin	0.10.1	0.10.1	0.10.1	0.10.1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

Notas da versão 6.13.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 6.13.0 da Amazon. As alterações são referentes à versão 6.12.0. Para obter informações sobre o cronograma da versão, consulte o [Log de alterações 6.13.0](#).

Novos atributos

- 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.

Alterações, melhorias e problemas resolvidos

- A versão 6.13.0 aprimora o daemon de gerenciamento de EMR registros da Amazon para garantir que todos os registros sejam carregados regularmente para o Amazon S3 quando um comando de encerramento de cluster é emitido. Isso facilita o encerramento mais rápido do cluster.
- A versão 6.13.0 aprimora os recursos de gerenciamento de logs EMR da Amazon para garantir o upload consistente e oportuno de todos os arquivos de log para o Amazon S3. Isso beneficia especialmente os EMR clusters de longa duração.
- Quando você inicia um cluster com a versão de patch mais recente do Amazon EMR 5.36 ou superior, 6.6 ou superior ou 7.0 ou superior, a Amazon EMR usa a versão mais recente do Amazon Linux 2023 ou Amazon Linux 2 para a Amazon padrão. EMR AMI Para obter mais informações, consulte [Usando o Amazon Linux padrão AMI para a Amazon EMR](#).

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 709.1	4.14.348	23 de julho de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			<p>Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia), Ásia Pacífico (Hyderabad), Oriente Médio (), Europa (EspanhaUAE), Europa (Zurique), Ásia-Pacífico (Melbourne), Israel (Tel Aviv), Oeste do Canadá (Calgary)</p>

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 223.0	4.14.336	8 de março de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelease Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia) , Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 131.0	4.14.336	14 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelease Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia), Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 124.0	4.14.336	7 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia), Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 109.0	4.14.334	24 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelease Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia) , Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 218.0	4.14.330	2 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 206.0	4.14.330	22 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelease Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 116.0	4.14.328	11 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 101.0	4.14.327	16 de novembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 020.1	4.14.326	7 de novembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 012.1	4.14.326	26 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 926.0	4.14.322	19 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 906.0	4.14.322	4 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 808.0	4.14.320	24 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)

Versões padrão do Java 6.13.0

As EMR versões 6.12.0 e superiores da Amazon oferecem suporte a todos os aplicativos com o Amazon Corretto 8 por padrão, exceto o Trino. Para Trino, a Amazon EMR oferece suporte ao Amazon Corretto 17 por padrão, começando com a versão 6.9.0 da EMR Amazon. A Amazon EMR também oferece suporte a alguns aplicativos com o Amazon Corretto 11 e 17. Essas aplicações estão listadas na tabela a seguir. Se você quiser alterar o padrão JVM em seu cluster, siga as [Configurar aplicações para usar uma Máquina Virtual Java específica](#) instruções em cada aplicativo executado no cluster. Você só pode usar uma versão de runtime do Java para um cluster. A Amazon EMR não oferece suporte à execução de diferentes nós ou aplicativos em diferentes versões de tempo de execução no mesmo cluster.

Embora a Amazon EMR ofereça suporte ao Amazon Corretto 11 e 17 no Apache Spark, no Apache Hadoop e no Apache Hive, o desempenho pode regredir em algumas cargas de trabalho quando você usa essas versões do Corretto. Recomendamos que você teste suas workloads antes de alterar os padrões.

A tabela a seguir mostra as versões padrão do Java para aplicativos na Amazon EMR 6.13.0:

Aplicativo	Versão Java/Amazon Corretto (o padrão é negrito)
Delta	17, 11, 8
Flink	11, 8
Ganglia	8
HBase	11, 8

Aplicativo	Versão Java/Amazon Corretto (o padrão é negrito)
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
Faísca RAPIDS	17, 11, 8
Sqoop	8
Tez	17, 11, 8
Trino	17
Zeppelin	8
Pig	8
Zookeeper	8

Versões do componente 6.13.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR

e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.4.2	Amazon SageMaker Spark SDK
<code>delta</code>	2.4.0	O Delta Lake é um formato de tabela aberto para grandes conjuntos de dados analíticos
<code>delta-standalone-connectors</code>	0.6.0	Os conectores do Delta oferecem runtimes diferentes para integrar o Delta Lake a mecanismos, como Flink, Hive e Presto.
<code>emr-ddb</code>	5.1.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	3.6.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.10.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-notebook-env</code>	1.7.0	Ambiente Conda para bloco de anotações do emr, que

Componente	Version (Versão)	Descrição
		inclui o jupyter enterprise gateway
emr-s3-dist-cp	2.27.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	2.6.0	EMRConector S3Select
emr-wal-cli	1.1.0	Cli usada para lista/exclusão de emrwal.
emrfs	2.58.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.17.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
flink-jobmanager-config	1.17.0	Gerenciando recursos em EMR nós para o Apache JobManager Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.

Componente	Version (Versão)	Descrição
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	3.3.3-amzn-5	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	3.3.3-amzn-5	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	3.3.3-amzn-5	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	3.3.3-amzn-5	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	3.3.3-amzn-5	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httpfs-server	3.3.3-amzn-5	HTTPendpoint para HDFS operações.
hadoop-kms-server	3.3.3-amzn-5	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	3.3.3-amzn-5	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.

Componente	Version (Versão)	Descrição
hadoop-yarn-nodemanager	3.3.3-amzn-5	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	3.3.3-amzn-5	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	3.3.3-amzn-5	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	2.4.17-amzn-1	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	2.4.17-amzn-1	Serviço para atender a uma ou mais HBase regiões.
hbase-client	2.4.17-amzn-1	HBasecliente de linha de comando.
hbase-rest-server	2.4.17-amzn-1	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	2.4.17-amzn-1	Serviço que fornece um endpoint Thrift para. HBase
hbase-operator-tools	2.4.17-amzn-1	Ferramenta de reparo para HBase clusters Apache.
hcatalog-client	3.1.3-amzn-6	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.

Componente	Version (Versão)	Descrição
hcatalog-server	3.1.3-amzn-6	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	3.1.3-amzn-6	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	3.1.3-amzn-6	O cliente da linha de comando do Hive.
hive-hbase	3.1.3-amzn-6	Cliente do Hive-hbase.
hive-metastore-server	3.1.3-amzn-6	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	3.1.3-amzn-6	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.13.1-amzn-1	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.
hudi-presto	0.13.1-amzn-1	Biblioteca de pacotes para executar o Presto com o Hudi.
hudi-trino	0.13.1-amzn-1	Biblioteca de pacotes para execução do Presto com o Hudi.

Componente	Version (Versão)	Descrição
hudi-spark	0.13.1-amzn-1	Biblioteca de pacotes para executar o Spark com o Hudi.
hue-server	4.11.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
iceberg	1.3.0-amzn-1	Apache Iceberg é um formato de tabela aberta para conjuntos de dados analíticos imensos
jupyterhub	1.5.0	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.7.1-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mxnet	1.9.1	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mariadb-server	5.5.68+	Servidor de banco de dados MariaDB.
nvidia-cuda	11.8.0	Drivers NVIDIA e toolkit CUDA
oozie-client	5.2.1	O cliente da linha de comando do Oozie.
oozie-server	5.2.1	O serviço que aceita solicitações de fluxo de trabalho do Oozie.

Componente	Version (Versão)	Descrição
opencv	4.7.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	5.1.3	As bibliotecas do phoenix para servidor e cliente
phoenix-connectors	5.1.3	Conectores do Apache Phoenix para Spark-3
phoenix-query-server	5.1.3	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.281-amzn-1	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.281-amzn-1	O serviço que executa partes de uma consulta.
presto-client	0.281-amzn-1	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
trino-coordinator	414-amzn-1	Serviço que aceita consultas e gerencia a execução de consultas entre trino-workers.
trino-worker	414-amzn-1	O serviço que executa partes de uma consulta.

Componente	Version (Versão)	Descrição
trino-client	414-amzn-1	Cliente de linha de comando do Trino instalado em mestres de espera de um cluster de alta disponibilidade (HA) em que o servidor do Trino não é iniciado.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	4.0.2	O projeto R para computação estatística
ranger-kms-server	2.0.0	Sistema de gerenciamento de chaves Apache Ranger
spark-client	3.4.1-amzn-0	Os clientes da linha de comando do Spark.
spark-history-server	3.4.1-amzn-0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	3.4.1-amzn-0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	3.4.1-amzn-0	Bibliotecas do Apache Spark necessárias para escravos. YARN
spark-rapids	23.06.0-amzn-1	RAPIDSPlugin Nvidia Spark que acelera o Apache Spark com o. GPUs

Componente	Version (Versão)	Descrição
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	2.11.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.10.2-amzn-4	O YARN aplicativo e as bibliotecas tez.
tez-on-worker	0.10.2-amzn-4	O YARN aplicativo tez e as bibliotecas para nós de trabalho.
webserver	2.4.41+	HTTPServidor Apache.
zeppelin-server	0.10.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.5.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.5.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 6.13.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Ocorrem ações de reconfiguração quando você especifica uma configuração para grupos de instâncias em um cluster em execução. A Amazon EMR só inicia ações de reconfiguração para as classificações que você modifica. Para ter mais informações, consulte [Reconfigurar um grupo de instâncias de um cluster em execução](#).

Classificações do emr-6.13.0

Classificações	Descrição	Ações de reconfiguração
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.	Restarts the ResourceManager service.
container-executor	Altere os valores no arquivo container-executor.cfg YARN do Hadoop.	Not available.
container-log4j	Altere os valores no arquivo container-log4j.properties YARN do Hadoop.	Not available.
core-site	Alterar os valores no arquivo core-site.xml do 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	Alterar as configurações relacionadas ao docker.	Not available.

Classificações	Descrição	Ações de reconfiguração
emrfs-site	Altere EMRFS as configurações.	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	Alterar as configurações do <code>flink-conf.yaml</code> .	Restarts Flink history server.
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.	Restarts Flink history server.
flink-log4j-session	Alterar as configurações de <code>log4j-yarn-session.properties</code> no Flink para sessão do Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Alterar as configurações de <code>log4j-cli.properties</code> no Flink.	Restarts Flink history server.

Classificações	Descrição	Ações de reconfiguração
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do 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	Alterar os valores no arquivo log4j.properties do 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	Alterar a configuração do servidor SSL no Hadoop	Not available.
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop	Not available.

Classificações	Descrição	Ações de reconfiguração
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Mude os valores no HBase ambiente.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.	Not available.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.	This classification should not be reconfigured.
hdfs-env	Mude os valores no HDFS ambiente.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classificações	Descrição	Ações de reconfiguração
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Mude os valores no HCatalog ambiente.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Configurações EMR selecionadas pela Amazon para o Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.	Not available.

Classificações	Descrição	Ações de reconfiguração
hive-parquet-logging	Alterar os valores no arquivo <code>parquet-logging.properties</code> do Hive.	Not available.
hive-env	Alterar os valores no ambiente do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive <code>schemaTool CLI</code> commands to verify <code>hive-metastore</code> .
hive-exec-log4j2	Altere os valores no arquivo <code>hive-exec-log 4j2.properties</code> do Hive.	Not available.
hive-llap-daemon-log4j2	Altere os valores no arquivo <code>llap-daemon-log 4j2.properties</code> do Hive.	Not available.
hive-log4j2	Alterar os valores no arquivo <code>hive-log4j2.properties</code> do Hive.	Not available.
hive-site	Alterar os valores no arquivo <code>hive-site.xml</code> do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive <code>schemaTool CLI</code> commands to verify <code>hive-metastore</code> . Also restarts Oozie and Zeppelin.
hiveserver2-site	Alterar os valores no arquivo <code>hiveserver2-site.xml</code> do Hive Server2.	Not available.
hue-ini	Alterar os valores no arquivo <code>ini</code> do Hue	Restarts Hue. Also activates Hue <code>config override CLI</code> commands to pick up new configurations.

Classificações	Descrição	Ações de reconfiguração
httpfs-env	Mude os valores no HTTPFS ambiente.	Restarts Hadoop Httpfs service.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.	Not available.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.	Restarts Hadoop-KMS service.
hadoop-kms-java-home	Mude a página inicial do Java do KMS Hadoop	Not available.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.	Not available.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Altere os valores no ambiente do Hudi.	Not available.
hudi-defaults	Alterar os valores no arquivo hudi-defaults.conf do Hudi.	Not available.
iceberg-defaults	Alterar os valores no arquivo iceberg-defaults.conf do Iceberg.	Not available.
delta-defaults	Alterar os valores no arquivo delta-defaults.conf do Delta.	Not available.

Classificações	Descrição	Ações de reconfiguração
jupyter-notebook-conf	Alterar os valores no arquivo <code>jupyter_notebook_config.py</code> do Notebook Jupyter.	Not available.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo <code>jupyterhub_config.py</code> do.	Not available.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.	Not available.
jupyter-sparkmagic-conf	Altere os valores no arquivo <code>config.json</code> do Sparkmagic.	Not available.
livy-conf	Alterar os valores no arquivo <code>livy.conf</code> do Livy.	Restarts Livy Server.
livy-env	Alterar os valores no ambiente do Livy.	Restarts Livy Server.
livy-log4j2	Alterar as configurações de <code>log4j2.properties</code> do Livy.	Restarts Livy Server.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Altere os valores no arquivo <code>mapred-site.xml</code> do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Alterar os valores no ambiente do Oozie.	Restarts Oozie.
oozie-log4j	Alterar os valores no arquivo <code>oozie-log4j.properties</code> do Oozie.	Restarts Oozie.

Classificações	Descrição	Ações de reconfiguração
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.	Not available.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.	Not available.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.	Not available.
pig-env	Alterar os valores no ambiente do Pig.	Not available.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.	Restarts Oozie.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.	Not available.
presto-log	Alterar os valores no arquivo log.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Alterar os valores no arquivo config.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-env	Altere os valores no arquivo <code>presto-env.sh</code> do Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Altere os valores no arquivo <code>node.properties</code> do Presto.	Not available.
presto-connector-blackhole	Alterar os valores no arquivo <code>blackhole.properties</code> do Presto.	Not available.
presto-connector-cassandra	Alterar os valores no arquivo <code>cassandra.properties</code> do Presto.	Not available.
presto-connector-hive	Alterar os valores no arquivo <code>hive.properties</code> do Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Alterar os valores no arquivo <code>jmx.properties</code> do Presto.	Not available.
presto-connector-kafka	Alterar os valores no arquivo <code>kafka.properties</code> do Presto.	Not available.
presto-connector-lakeformat ion	Alterar os valores no arquivo <code>lakeformation.properties</code> do Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Alterar os valores no arquivo <code>localfile.properties</code> do Presto.	Not available.
presto-connector-memory	Alterar os valores no arquivo <code>memory.properties</code> do Presto.	Not available.
presto-connector-mongodb	Alterar os valores no arquivo <code>mongodb.properties</code> do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.	Not available.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.	Not available.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.	Not available.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.	Not available.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.	Not available.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.	Not available.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.	Not available.
trino-log	Alterar os valores no arquivo log.properties do Trino.	Restarts Trino-Server (for Trino)
trino-config	Alterar os valores no arquivo config.properties do Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Alterar os valores no arquivo password-authenticator.properties do Trino.	Restarts Trino-Server (for Trino)
trino-env	Alterar os valores no arquivo trino-env.sh do Trino.	Restarts Trino-Server (for Trino)

Classificações	Descrição	Ações de reconfiguração
trino-node	Alterar os valores no arquivo <code>node.properties</code> do Trino.	Not available.
trino-connector-blackhole	Alterar os valores no arquivo <code>blackhole.properties</code> do Trino.	Not available.
trino-connector-cassandra	Alterar os valores no arquivo <code>cassandra.properties</code> do Trino.	Not available.
trino-connector-delta	Alterar os valores no arquivo <code>delta.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Alterar os valores no arquivo <code>hive.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Alterar os valores no arquivo <code>exchange-manager.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Alterar os valores no arquivo <code>iceberg.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Alterar os valores no arquivo <code>hudi.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Alterar os valores no arquivo <code>jmx.properties</code> do Trino.	Not available.
trino-connector-kafka	Alterar os valores no arquivo <code>kafka.properties</code> do Trino.	Not available.
trino-connector-localfile	Alterar os valores no arquivo <code>localfile.properties</code> do Trino.	Not available.
trino-connector-memory	Alterar os valores no arquivo <code>memory.properties</code> do Trino.	Not available.

Classificações	Descrição	Ações de reconfiguração
trino-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Trino.	Not available.
trino-connector-mysql	Alterar os valores no arquivo mysql.properties do Trino.	Not available.
trino-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Trino.	Not available.
trino-connector-raptor	Alterar os valores no arquivo raptor.properties do Trino.	Not available.
trino-connector-redis	Alterar os valores no arquivo redis.properties do Trino.	Not available.
trino-connector-redshift	Alterar os valores no arquivo redshift.properties do Trino.	Not available.
trino-connector-tpch	Alterar os valores no arquivo tpch.properties do Trino.	Not available.
trino-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Trino.	Not available.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.	Restarts Ranger KMS Server.
ranger-kms-logback	Altere os valores no arquivo kms-logback.xml do Ranger. KMS	Not available.

Classificações	Descrição	Ações de reconfiguração
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.	Not available.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Alterar os valores no ambiente do Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.	Not available.
spark-log4j2	Alterar os valores no arquivo log4j2.properties do Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Alterar os valores no ambiente do Sqoop.	Not available.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.	Not available.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.	Not available.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.	Restart Oozie and HiveServer2.

Classificações	Descrição	Ações de reconfiguração
yarn-env	Mude os valores no YARN ambiente.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.	Restarts Zeppelin.
zeppelin-site	Alterar as definições da configuração em zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.	Restarts Zookeeper server.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.	Restarts Zookeeper server.

Log de alterações 6.13.0

Log de alterações da versão 6.13.0 e das notas da versão

Data	Evento	Descrição
2023-09-23	Implantação concluída	Amazon EMR 6.13.0 totalmente implantado em todas as regiões suportadas
2023-09-12	Publicação de documentos	Notas de lançamento do Amazon EMR 6.13.0 publicadas pela primeira vez
2023-09-01	Lançamento inicial	Amazon EMR 6.13.0 foi implantado pela primeira vez nas regiões comerciais iniciais

Amazon EMR versão 6.12.0

Versões da aplicação 6.12.0

As seguintes aplicações são compatíveis com esta versão: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-6.12.0	emr-6.11.1	emr-6.11.0	emr-6.10.1
AWS SDK para 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	2.12.15	2.12.15	2.12.15	2.12.15
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
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

	emr-6.12.0	emr-6.11.1	emr-6.11.0	emr-6.10.1
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	0.10.2	0.10.2	0.10.2	0.10.2
Trino (PrestoSQL)	414	410	410	403
Zeppelin	0.10.1	0.10.1	0.10.1	0.10.1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

Notas da versão 6.12.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 6.12.0 da Amazon. As alterações são referentes à versão 6.11.0. Para obter informações sobre o cronograma da versão, consulte o [Log de alterações 6.12.0](#).

Novos atributos

- 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.
- A Amazon EMR lança 6.12.0 e versões superiores de suporte à LDAP integração com Apache Livy, Apache Hive through HiveServer 2 (HS2), Trino, Presto e Hue. Você também pode instalar

o Apache Spark e o Apache Hadoop em um EMR cluster que usa 6.12.0 ou superior e configurá-los para uso. LDAP Para obter mais informações, consulte [Usar o Active Directory ou LDAP servidores para autenticação com a Amazon EMR](#).

Alterações, melhorias e problemas resolvidos

- As EMR versões 6.12.0 e superiores da Amazon oferecem suporte ao tempo de execução do Java 11 para o Flink. Para ter mais informações, consulte [Configurar o Flink para ser executado com o Java 11](#).
- A versão 6.12.0 adiciona um novo mecanismo de repetição ao fluxo de trabalho de escalabilidade de EMR clusters para clusters que executam o Presto ou o Trino. Essa melhoria reduz o risco de que o redimensionamento do cluster fique paralisado indefinidamente devido a uma única falha na operação de redimensionamento. Ela também aprimora a utilização dos clusters, porque seu cluster aumenta e reduz a escala verticalmente com mais rapidez.
- A versão 6.12.0 corrige um problema em que as operações de redução da escala verticalmente do cluster podem ficar paralisadas quando um nó central que está passando por uma desativação tranquila se torna não íntegro por qualquer motivo antes de ser totalmente desativado.
- A versão 6.12.0 melhora a lógica de redução do cluster para que seu cluster não tente reduzir os nós principais abaixo da configuração do fator de HDFS replicação do cluster. Isso se alinha aos seus requisitos de redundância de dados e reduz a probabilidade de uma operação de escalabilidade paralisar.
- A versão 6.12.0 aprimora o desempenho e a eficiência do serviço de monitoramento de saúde da Amazon, EMR aumentando a velocidade com que ele registra as mudanças de estado das instâncias. Essa melhoria reduz a probabilidade de degradação do desempenho dos nós do cluster que estão executando várias ferramentas de cliente ou aplicações de terceiros personalizadas.
- A versão 6.12.0 melhora o desempenho do daemon de gerenciamento de logs no cluster para a Amazon. EMR Como resultado, há menos chances de degradação do desempenho com EMR clusters que executam etapas com alta simultaneidade.
- Com a EMR versão 6.12.0 da Amazon, o daemon de gerenciamento de registros foi atualizado para identificar todos os registros que estão em uso ativo com identificadores de arquivos abertos no armazenamento da instância local e os processos associados. Essa atualização garante que a Amazon exclua EMR adequadamente os arquivos e recupere o espaço de armazenamento depois que os registros forem arquivados no Amazon S3.

- A versão 6.12.0 inclui um aprimoramento do daemon de gerenciamento de logs que exclui diretórios de etapas vazios e não utilizados no sistema de arquivos de cluster local. Um número excessivamente grande de diretórios vazios pode degradar o desempenho dos EMR daemons da Amazon e resultar na sobreutilização do disco.
- A versão 6.12.0 permite a rotação de registros para registros YARN do Timeline Server. Isso minimiza os cenários de utilização excessiva do disco, especialmente para clusters de execução prolongada.
- O tamanho padrão do volume raiz aumentou para 15 GB no Amazon EMR 6.10.0 e superior. O tamanho padrão do volume raiz das versões anteriores é de 10 GB.
- Quando você inicia um cluster com a versão de patch mais recente do Amazon EMR 5.36 ou superior, 6.6 ou superior ou 7.0 ou superior, a Amazon EMR usa a versão mais recente do Amazon Linux 2023 ou Amazon Linux 2 para a Amazon padrão. EMR AMI Para obter mais informações, consulte [Usando o Amazon Linux padrão AMI para a Amazon EMR](#).

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 709.1	4.14.348	23 de julho de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			<p>Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia), Ásia Pacífico (Hyderabad), Oriente Médio (), Europa (EspanhaUAE), Europa (Zurique), Ásia-Pacífico (Melbourne), Israel (Tel Aviv), Oeste do Canadá (Calgary)</p>

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 223.0	4.14.336	8 de março de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia), Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 131.0	4.14.336	14 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelease Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia) , Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 124.0	4.14.336	7 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia) , Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 109.0	4.14.334	24 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia) , Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 218.0	4.14.330	2 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (Doha), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 206.0	4.14.330	22 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 116.0	4.14.328	11 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 101.0	4.14.327	16 de novembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 020.1	4.14.326	7 de novembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 012.1	4.14.326	26 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelease Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 926.0	4.14.322	19 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 906.0	4.14.322	4 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 822.0	4.14.322	30 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 808.0	4.14.320	24 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 727.0	4.14.320	14 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 719.0	4.14.320	2 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 628.0	4.14.318	12 de julho de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Oriente Médio (), Canadá (Central) UAE

Versões padrão do Java 6.12.0

As EMR versões 6.12.0 e superiores da Amazon oferecem suporte a todos os aplicativos com o Amazon Corretto 8 por padrão, exceto o Trino. Para Trino, a Amazon EMR oferece suporte ao Amazon Corretto 17 por padrão, começando com a versão 6.9.0 da EMR Amazon. A Amazon EMR também oferece suporte a alguns aplicativos com o Amazon Corretto 11 e 17. Essas aplicações estão listadas na tabela a seguir. Se você quiser alterar o padrão JVM em seu cluster, siga as [Configurar aplicações para usar uma Máquina Virtual Java específica](#) instruções em cada aplicativo executado no cluster. Você só pode usar uma versão de runtime do Java para um cluster. A Amazon EMR não oferece suporte à execução de diferentes nós ou aplicativos em diferentes versões de tempo de execução no mesmo cluster.

Embora a Amazon EMR ofereça suporte ao Amazon Corretto 11 e 17 no Apache Spark, no Apache Hadoop e no Apache Hive, o desempenho pode regredir em algumas cargas de trabalho quando você usa essas versões do Corretto. Recomendamos que você teste suas workloads antes de alterar os padrões.

A tabela a seguir mostra as versões padrão do Java para aplicativos na Amazon EMR 6.12.0:

Aplicativo	Versão Java/Amazon Corretto (o padrão é negrito)
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
Iceberg	17, 11, 8

Aplicativo	Versão Java/Amazon Corretto (o padrão é negrito)
Livy	17, 11, 8
Oozie	17, 11, 8
Phoenix	8
PrestoDB	8
Spark	17, 11, 8
Faísca RAPIDS	17, 11, 8
Sqoop	8
Tez	17, 11, 8
Trino	17
Zeppelin	8
Pig	8
Zookeeper	8

Versões de componente 6.12.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto

nomeado myapp-component com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como 2.2-amzn-2.

Componente	Version (Versão)	Descrição
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark SDK
delta	2.4.0	O Delta Lake é um formato de tabela aberto para grandes conjuntos de dados analíticos
delta-standalone-connectors	0.6.0	Os conectores do Delta oferecem runtimes diferentes para integrar o Delta Lake a mecanismos, como Flink, Hive e Presto.
emr-ddb	5.1.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	3.5.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.9.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-notebook-env	1.7.0	Ambiente Conda para bloco de anotações do emr, que inclui o jupyter enterprise gateway
emr-s3-dist-cp	2.26.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	2.5.0	EMRConector S3Select

Componente	Version (Versão)	Descrição
emr-wal-cli	1.1.0	Cli usada para lista/exclusão de emrwal.
emrfs	2.57.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.17.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
flink-jobmanager-config	1.17.0	Gerenciando recursos em EMR nós para o Apache JobManager Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	3.3.3-amzn-4	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	3.3.3-amzn-4	HDFSserviço em nível de nó para armazenar blocos.

Componente	Version (Versão)	Descrição
hadoop-hdfs-library	3.3.3-amzn-4	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	3.3.3-amzn-4	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	3.3.3-amzn-4	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-https-server	3.3.3-amzn-4	HTTPendpoint para HDFS operações.
hadoop-kms-server	3.3.3-amzn-4	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	3.3.3-amzn-4	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	3.3.3-amzn-4	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	3.3.3-amzn-4	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	3.3.3-amzn-4	Serviço para recuperar informações atuais e históricas para YARN aplicativos.

Componente	Version (Versão)	Descrição
hbase-hmaster	2.4.17-amzn-0	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	2.4.17-amzn-0	Serviço para atender a uma ou mais HBase regiões.
hbase-client	2.4.17-amzn-0	HBase cliente de linha de comando.
hbase-rest-server	2.4.17-amzn-0	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	2.4.17-amzn-0	Serviço que fornece um endpoint Thrift para HBase.
hbase-operator-tools	2.4.17-amzn-0	Ferramenta de reparo para HBase clusters Apache.
hcatalog-client	3.1.3-amzn-5	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	3.1.3-amzn-5	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	3.1.3-amzn-5	HTTP endpoint fornecendo uma REST interface para HCatalog.

Componente	Version (Versão)	Descrição
hive-client	3.1.3-amzn-5	O cliente da linha de comando do Hive.
hive-hbase	3.1.3-amzn-5	Cliente do Hive-hbase.
hive-metastore-server	3.1.3-amzn-5	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	3.1.3-amzn-5	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.13.1-amzn-0	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.
hudi-presto	0.13.1-amzn-0	Biblioteca de pacotes para executar o Presto com o Hudi.
hudi-trino	0.13.1-amzn-0	Biblioteca de pacotes para execução do Presto com o Hudi.
hudi-spark	0.13.1-amzn-0	Biblioteca de pacotes para executar o Spark com o Hudi.
hue-server	4.11.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop

Componente	Version (Versão)	Descrição
iceberg	1.3.0-amzn-0	Apache Iceberg é um formato de tabela aberta para conjuntos de dados analíticos imensos
jupyterhub	1.4.1	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.7.1-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mxnet	1.9.1	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mariadb-server	5.5.68+	Servidor de banco de dados MariaDB.
nvidia-cuda	11.8.0	Drivers NVIDIA e toolkit CUDA
oozie-client	5.2.1	O cliente da linha de comando do Oozie.
oozie-server	5.2.1	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	4.7.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	5.1.3	As bibliotecas do phoenix para servidor e cliente

Componente	Version (Versão)	Descrição
phoenix-connectors	5.1.3	Conectores do Apache Phoenix para Spark-3
phoenix-query-server	5.1.3	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.281-amzn-0	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.281-amzn-0	O serviço que executa partes de uma consulta.
presto-client	0.281-amzn-0	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
trino-coordinator	414-amzn-0	Serviço que aceita consultas e gerencia a execução de consultas entre trino-workers.
trino-worker	414-amzn-0	O serviço que executa partes de uma consulta.
trino-client	414-amzn-0	Cliente de linha de comando do Trino instalado em mestres de espera de um cluster de alta disponibilidade (HA) em que o servidor do Trino não é iniciado.

Componente	Version (Versão)	Descrição
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	4.0.2	O projeto R para computação estatística
ranger-kms-server	2.0.0	Sistema de gerenciamento de chaves Apache Ranger
spark-client	3.4.0-amzn-0	Os clientes da linha de comando do Spark.
spark-history-server	3.4.0-amzn-0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	3.4.0-amzn-0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	3.4.0-amzn-0	Bibliotecas do Apache Spark necessárias para escravos. YARN
spark-rapids	23.06.0-amzn-0	RAPIDSPlugin Nvidia Spark que acelera o Apache Spark com o. GPUs
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	2.11.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.

Componente	Version (Versão)	Descrição
tez-on-yarn	0.10.2-amzn-3	O YARN aplicativo e as bibliotecas tez.
tez-on-worker	0.10.2-amzn-3	O YARN aplicativo tez e as bibliotecas para nós de trabalho.
webserver	2.4.41+	HTTPServidor Apache.
zeppelin-server	0.10.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.5.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.5.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 6.12.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Ocorrem ações de reconfiguração quando você especifica uma configuração para grupos de instâncias em um cluster em execução. A Amazon EMR só inicia ações de reconfiguração para as classificações que você modifica. Para ter mais informações, consulte [Reconfigurar um grupo de instâncias de um cluster em execução](#).

Classificações do emr-6.12.0

Classificações	Descrição	Ações de reconfiguração
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.	Restarts the ResourceManager service.
container-executor	Altere os valores no arquivo container-executor.cfg YARN do Hadoop.	Not available.
container-log4j	Altere os valores no arquivo container-log4j.properties YARN do Hadoop.	Not available.
core-site	Alterar os valores no arquivo core-site.xml do 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	Alterar as configurações relacionadas ao docker.	Not available.
emrfs-site	Altere EMRFS as configurações.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

Classificações	Descrição	Ações de reconfiguração
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Alterar as configurações do flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Alterar as configurações de log4j.properties no Flink.	Restarts Flink history server.
flink-log4j-session	Alterar as configurações de log4j-yarn-session.properties no Flink para sessão do Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.	Restarts Flink history server.

Classificações	Descrição	Ações de reconfiguração
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do 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	Alterar os valores no arquivo log4j.properties do 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	Alterar a configuração do servidor SSL no Hadoop	Not available.
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop	Not available.

Classificações	Descrição	Ações de reconfiguração
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Mude os valores no HBase ambiente.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.	Not available.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.	This classification should not be reconfigured.
hdfs-env	Mude os valores no HDFS ambiente.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classificações	Descrição	Ações de reconfiguração
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Mude os valores no HCatalog ambiente.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Configurações EMR selecionadas pela Amazon para o Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.	Not available.

Classificações	Descrição	Ações de reconfiguração
hive-parquet-logging	Alterar os valores no arquivo <code>parquet-logging.properties</code> do Hive.	Not available.
hive-env	Alterar os valores no ambiente do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive <code>schemaTool CLI</code> commands to verify <code>hive-metastore</code> .
hive-exec-log4j2	Altere os valores no arquivo <code>hive-exec-log 4j2.properties</code> do Hive.	Not available.
hive-llap-daemon-log4j2	Altere os valores no arquivo <code>llap-daemon-log 4j2.properties</code> do Hive.	Not available.
hive-log4j2	Alterar os valores no arquivo <code>hive-log4j2.properties</code> do Hive.	Not available.
hive-site	Alterar os valores no arquivo <code>hive-site.xml</code> do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive <code>schemaTool CLI</code> commands to verify <code>hive-metastore</code> . Also restarts Oozie and Zeppelin.
hiveserver2-site	Alterar os valores no arquivo <code>hiveserver2-site.xml</code> do Hive Server2.	Not available.
hue-ini	Alterar os valores no arquivo <code>ini</code> do Hue	Restarts Hue. Also activates Hue <code>config override CLI</code> commands to pick up new configurations.

Classificações	Descrição	Ações de reconfiguração
httpfs-env	Mude os valores no HTTPFS ambiente.	Restarts Hadoop Httpfs service.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.	Not available.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.	Restarts Hadoop-KMS service.
hadoop-kms-java-home	Mude a página inicial do Java do KMS Hadoop	Not available.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.	Not available.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Altere os valores no ambiente do Hudi.	Not available.
hudi-defaults	Alterar os valores no arquivo hudi-defaults.conf do Hudi.	Not available.
iceberg-defaults	Alterar os valores no arquivo iceberg-defaults.conf do Iceberg.	Not available.
delta-defaults	Alterar os valores no arquivo delta-defaults.conf do Delta.	Not available.

Classificações	Descrição	Ações de reconfiguração
jupyter-notebook-conf	Alterar os valores no arquivo <code>jupyter_notebook_config.py</code> do Notebook Jupyter.	Not available.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo <code>jupyterhub_config.py</code> do.	Not available.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.	Not available.
jupyter-sparkmagic-conf	Altere os valores no arquivo <code>config.json</code> do Sparkmagic.	Not available.
livy-conf	Alterar os valores no arquivo <code>livy.conf</code> do Livy.	Restarts Livy Server.
livy-env	Alterar os valores no ambiente do Livy.	Restarts Livy Server.
livy-log4j2	Alterar as configurações de <code>log4j2.properties</code> do Livy.	Restarts Livy Server.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Altere os valores no arquivo <code>mapred-site.xml</code> do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Alterar os valores no ambiente do Oozie.	Restarts Oozie.
oozie-log4j	Alterar os valores no arquivo <code>oozie-log4j.properties</code> do Oozie.	Restarts Oozie.

Classificações	Descrição	Ações de reconfiguração
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.	Not available.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.	Not available.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.	Not available.
pig-env	Alterar os valores no ambiente do Pig.	Not available.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.	Restarts Oozie.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.	Not available.
presto-log	Alterar os valores no arquivo log.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Alterar os valores no arquivo config.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-env	Altere os valores no arquivo presto-env.sh do Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Altere os valores no arquivo node.properties do Presto.	Not available.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.	Not available.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.	Not available.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.	Not available.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.	Not available.
presto-connector-lakeformat ion	Alterar os valores no arquivo lakeformation.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.	Not available.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.	Not available.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.	Not available.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.	Not available.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.	Not available.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.	Not available.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.	Not available.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.	Not available.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.	Not available.
trino-log	Alterar os valores no arquivo log.properties do Trino.	Restarts Trino-Server (for Trino)
trino-config	Alterar os valores no arquivo config.properties do Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Alterar os valores no arquivo password-authenticator.properties do Trino.	Restarts Trino-Server (for Trino)
trino-env	Alterar os valores no arquivo trino-env.sh do Trino.	Restarts Trino-Server (for Trino)

Classificações	Descrição	Ações de reconfiguração
trino-node	Alterar os valores no arquivo <code>node.properties</code> do Trino.	Not available.
trino-connector-blackhole	Alterar os valores no arquivo <code>blackhole.properties</code> do Trino.	Not available.
trino-connector-cassandra	Alterar os valores no arquivo <code>cassandra.properties</code> do Trino.	Not available.
trino-connector-delta	Alterar os valores no arquivo <code>delta.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Alterar os valores no arquivo <code>hive.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Alterar os valores no arquivo <code>exchange-manager.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Alterar os valores no arquivo <code>iceberg.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Alterar os valores no arquivo <code>hudi.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Alterar os valores no arquivo <code>jmx.properties</code> do Trino.	Not available.
trino-connector-kafka	Alterar os valores no arquivo <code>kafka.properties</code> do Trino.	Not available.
trino-connector-localfile	Alterar os valores no arquivo <code>localfile.properties</code> do Trino.	Not available.
trino-connector-memory	Alterar os valores no arquivo <code>memory.properties</code> do Trino.	Not available.

Classificações	Descrição	Ações de reconfiguração
trino-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Trino.	Not available.
trino-connector-mysql	Alterar os valores no arquivo mysql.properties do Trino.	Not available.
trino-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Trino.	Not available.
trino-connector-raptor	Alterar os valores no arquivo raptor.properties do Trino.	Not available.
trino-connector-redis	Alterar os valores no arquivo redis.properties do Trino.	Not available.
trino-connector-redshift	Alterar os valores no arquivo redshift.properties do Trino.	Not available.
trino-connector-tpch	Alterar os valores no arquivo tpch.properties do Trino.	Not available.
trino-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Trino.	Not available.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.	Restarts Ranger KMS Server.
ranger-kms-logback	Altere os valores no arquivo kms-logback.xml do Ranger. KMS	Not available.

Classificações	Descrição	Ações de reconfiguração
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.	Not available.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Alterar os valores no ambiente do Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.	Not available.
spark-log4j2	Alterar os valores no arquivo log4j2.properties do Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Alterar os valores no ambiente do Sqoop.	Not available.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.	Not available.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.	Not available.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.	Restart Oozie and HiveServer2.

Classificações	Descrição	Ações de reconfiguração
yarn-env	Mude os valores no YARN ambiente.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.	Restarts Zeppelin.
zeppelin-site	Alterar as definições da configuração em zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.	Restarts Zookeeper server.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.	Restarts Zookeeper server.

Log de alterações 6.12.0

Log de alterações da versão 6.12.0 e das notas da versão

Data	Evento	Descrição
2023-07-27	Atualizar a documentação	Atualize as opções Java para 6.12 e adicione o tutorial Oozie para atualizar JVM
2023-07-21	Implantação concluída	Amazon EMR 6.12.0 totalmente implantado em todas as regiões suportadas
2023-07-21	Publicação de documentos	Notas de lançamento do Amazon EMR 6.12.0 publicadas pela primeira vez
2023-07-12	Lançamento inicial	Amazon EMR 6.12.0 foi implantado pela primeira vez nas regiões comerciais iniciais

Amazon EMR versão 6.11.1

Versões 6.11.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)

- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-6.11.1	emr-6.11.0	emr-6.10.1	emr-6.10.0
AWS SDK para 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	2.12.15	2.12.15	2.12.15	2.12.15
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
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

	emr-6.11.1	emr-6.11.0	emr-6.10.1	emr-6.10.0
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	0.10.2	0.10.2	0.10.2	0.10.2
Trino (PrestoSQL)	410	410	403	403
Zeppelin	0.10.1	0.10.1	0.10.1	0.10.1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

Notas da versão 6.11.1

As notas de lançamento a seguir incluem informações sobre a EMR versão 6.11.1 da Amazon. As alterações são referentes à versão 6.11.0. Para obter informações sobre o cronograma da versão, consulte o [Log de alterações 6.11.1](#).

Alterações, melhorias e problemas resolvidos

- Devido à contenção de bloqueio, um nó pode entrar em um deadlock se for adicionado ou removido ao mesmo tempo em que tenta ser desativado. Como resultado, o Hadoop Resource

Manager (YARN) deixa de responder e afeta todos os contêineres recebidos e em execução no momento.

- Esta versão inclui uma alteração que permite que clusters de alta disponibilidade se recuperem de um estado de falha após a reinicialização.
- Esta versão inclui correções de segurança para Hue e HBase
- Esta versão corrige um problema em que clusters que estão executando cargas de trabalho no Spark com a Amazon EMR podem receber silenciosamente resultados incorretos com `contains`, `startsWith`, e `endsWith` `like`. Esse problema ocorre quando você usa as expressões em campos particionados que têm metadados no Amazon EMR Hive3 Metastore Server (). HMS
- Esta versão corrige um problema de limitação no lado do Glue quando não há funções definidas pelo usuário (). UDF
- Esta versão corrige um problema que exclui registros de contêineres pelo serviço de agregação de registros de nós antes que o carregador de registros possa enviá-los para o S3 em caso de descomissionamento. YARN
- Esta versão corrige um problema com as métricas do FairShare Scheduler quando o Node Label está habilitado para o Hadoop.
- Esta versão corrige um problema que afetou a performance do Spark quando você definiu um valor de `true` padrão para a configuração `spark.yarn.heterogeneousExecutors.enabled` no `spark-defaults.conf`.
- Esta versão corrige um problema com a falha do Reduce Task em ler dados embaralhados. O problema causou falhas na consulta do Hive com um erro de memória corrompida.
- Esta versão adiciona um novo mecanismo de repetição ao fluxo de trabalho de escalabilidade de EMR clusters para clusters que executam o Presto ou o Trino. Essa melhoria reduz o risco de que o redimensionamento do cluster fique paralisado indefinidamente devido a uma única falha na operação de redimensionamento. Ela também aprimora a utilização dos clusters, porque seu cluster aumenta e reduz a escala verticalmente com mais rapidez.
- Esta versão melhora a lógica de redução do cluster para que seu cluster não tente reduzir os nós principais abaixo da configuração do fator de HDFS replicação do cluster. Isso se alinha aos seus requisitos de redundância de dados e reduz a probabilidade de uma operação de escalabilidade paralisar.
- O daemon de gerenciamento de logs foi atualizado para identificar todos os logs que estão em uso ativo com identificadores de arquivos abertos no armazenamento da instância local e nos processos associados. Essa atualização garante que a Amazon exclua EMR adequadamente os

arquivos e recupere o espaço de armazenamento depois que os registros forem arquivados no Amazon S3.

- Esta versão inclui um aprimoramento do daemon de gerenciamento de logs que exclui diretórios de etapas vazios e não utilizados no sistema de arquivos de cluster local. Um número excessivamente grande de diretórios vazios pode degradar o desempenho dos EMR daemons da Amazon e resultar na sobreutilização do disco.
- Quando você inicia um cluster com a versão de patch mais recente do Amazon EMR 5.36 ou superior, 6.6 ou superior ou 7.0 ou superior, a Amazon EMR usa a versão mais recente do Amazon Linux 2023 ou Amazon Linux 2 para a Amazon padrão. EMR AMI Para obter mais informações, consulte [Usando o Amazon Linux padrão AMI para a Amazon EMR](#).

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 709.1	4.14.348	23 de julho de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia), Ásia Pacífico (Hyderabad), Oriente Médio (), Europa (EspanhaUAE), Europa (Zurique), Ásia-Pacífico (Melbourne), Israel (Tel Aviv), Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 223.0	4.14.336	8 de março de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia), Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 131.0	4.14.336	14 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia) , Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 124.0	4.14.336	7 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelease Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia) , Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 109.0	4.14.334	24 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia), Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 218.0	4.14.330	2 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 206.0	4.14.330	22 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 116.0	4.14.328	11 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 101.0	4.14.327	16 de novembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 020.1	4.14.326	7 de novembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 012.1	4.14.326	26 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 926.0	4.14.322	19 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 906.0	4.14.322	4 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 822.0	4.14.322	30 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 808.0	4.14.320	24 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)
2.0.2023 727.0	4.14.320	14 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central)

Versões de componente 6.11.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR

e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.4.2	Amazon SageMaker Spark SDK
<code>delta</code>	2.2.0	O Delta Lake é um formato de tabela aberto para grandes conjuntos de dados analíticos
<code>delta-standalone-connectors</code>	0.6.0	Os conectores do Delta oferecem runtimes diferentes para integrar o Delta Lake a mecanismos, como Flink, Hive e Presto.
<code>emr-ddb</code>	5.1.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	3.4.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.8.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-notebook-env</code>	1.7.0	Ambiente Conda para bloco de anotações do emr, que

Componente	Version (Versão)	Descrição
		inclui o jupyter enterprise gateway
emr-s3-dist-cp	2.25.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	2.4.0	EMRConector S3Select
emr-wal-cli	1.1.0	Cli usada para lista/exclusão de emrwal.
emrfs	2.56.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.16.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
flink-jobmanager-config	1.16.0	Gerenciando recursos em EMR nós para o Apache JobManager Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.

Componente	Version (Versão)	Descrição
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	3.3.3-amzn-3.1	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	3.3.3-amzn-3.1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	3.3.3-amzn-3.1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	3.3.3-amzn-3.1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	3.3.3-amzn-3.1	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httpfs-server	3.3.3-amzn-3.1	HTTPendpoint para HDFS operações.
hadoop-kms-server	3.3.3-amzn-3.1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	3.3.3-amzn-3.1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.

Componente	Version (Versão)	Descrição
hadoop-yarn-nodemanager	3.3.3-amzn-3.1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	3.3.3-amzn-3.1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	3.3.3-amzn-3.1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	2.4.15-amzn-1.1	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	2.4.15-amzn-1.1	Serviço para atender a uma ou mais HBase regiões.
hbase-client	2.4.15-amzn-1.1	HBasecliente de linha de comando.
hbase-rest-server	2.4.15-amzn-1.1	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	2.4.15-amzn-1.1	Serviço que fornece um endpoint Thrift para. HBase
hbase-operator-tools	2.4.15-amzn-1.1	Ferramenta de reparo para HBase clusters Apache.
hcatalog-client	3.1.3-amzn-4.1	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.

Componente	Version (Versão)	Descrição
hcatalog-server	3.1.3-amzn-4.1	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	3.1.3-amzn-4.1	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	3.1.3-amzn-4.1	O cliente da linha de comando do Hive.
hive-hbase	3.1.3-amzn-4.1	Cliente do Hive-hbase.
hive-metastore-server	3.1.3-amzn-4.1	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	3.1.3-amzn-4.1	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.13.0-amzn-0	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.
hudi-presto	0.13.0-amzn-0	Biblioteca de pacotes para executar o Presto com o Hudi.
hudi-trino	0.13.0-amzn-0	Biblioteca de pacotes para execução do Presto com o Hudi.

Componente	Version (Versão)	Descrição
hudi-spark	0.13.0-amzn-0	Biblioteca de pacotes para executar o Spark com o Hudi.
hue-server	4.11.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
iceberg	1.2.0-amzn-0	Apache Iceberg é um formato de tabela aberta para conjuntos de dados analíticos imensos
jupyterhub	1.4.1	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.7.1-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mxnet	1.9.1	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mariadb-server	5.5.68+	Servidor de banco de dados MariaDB.
nvidia-cuda	11.8.0	Drivers NVIDIA e toolkit CUDA
oozie-client	5.2.1	O cliente da linha de comando do Oozie.
oozie-server	5.2.1	O serviço que aceita solicitações de fluxo de trabalho do Oozie.

Componente	Version (Versão)	Descrição
opencv	4.5.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	5.1.2	As bibliotecas do phoenix para servidor e cliente
phoenix-connectors	5.1.2	Conectores do Apache Phoenix para Spark-3
phoenix-query-server	5.1.2	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.279-amzn-0	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.279-amzn-0	O serviço que executa partes de uma consulta.
presto-client	0.279-amzn-0	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
trino-coordinator	410-amzn-0	Serviço que aceita consultas e gerencia a execução de consultas entre trino-workers.
trino-worker	410-amzn-0	O serviço que executa partes de uma consulta.

Componente	Version (Versão)	Descrição
trino-client	410-amzn-0	Cliente de linha de comando do Trino instalado em mestres de espera de um cluster de alta disponibilidade (HA) em que o servidor do Trino não é iniciado.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	4.0.2	O projeto R para computação estatística
ranger-kms-server	2.0.0	Sistema de gerenciamento de chaves Apache Ranger
spark-client	3.3.2-amzn-0.1	Os clientes da linha de comando do Spark.
spark-history-server	3.3.2-amzn-0.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	3.3.2-amzn-0.1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	3.3.2-amzn-0.1	Bibliotecas do Apache Spark necessárias para escravos. YARN
spark-rapids	23.02.0-amzn-0	RAPIDSPlugin Nvidia Spark que acelera o Apache Spark com o. GPUs

Componente	Version (Versão)	Descrição
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	2.11.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.10.2-amzn-2.1	O YARN aplicativo e as bibliotecas tez.
tez-on-worker	0.10.2-amzn-2.1	O YARN aplicativo tez e as bibliotecas para nós de trabalho.
webserver	2.4.41+	HTTPServidor Apache.
zeppelin-server	0.10.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.5.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.5.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 6.11.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Ocorrem ações de reconfiguração quando você especifica uma configuração para grupos de instâncias em um cluster em execução. A Amazon EMR só inicia ações de reconfiguração para as classificações que você modifica. Para ter mais informações, consulte [Reconfigurar um grupo de instâncias de um cluster em execução](#).

Classificações do emr-6.11.1

Classificações	Descrição	Ações de reconfiguração
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.	Restarts the ResourceManager service.
container-executor	Altere os valores no arquivo container-executor.cfg YARN do Hadoop.	Not available.
container-log4j	Altere os valores no arquivo container-log4j.properties YARN do Hadoop.	Not available.
core-site	Alterar os valores no arquivo core-site.xml do 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	Alterar as configurações relacionadas ao docker.	Not available.

Classificações	Descrição	Ações de reconfiguração
emrfs-site	Altere EMRFS as configurações.	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	Alterar as configurações do <code>flink-conf.yaml</code> .	Restarts Flink history server.
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.	Restarts Flink history server.
flink-log4j-session	Alterar as configurações de <code>log4j-yarn-session.properties</code> no Flink para sessão do Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Alterar as configurações de <code>log4j-cli.properties</code> no Flink.	Restarts Flink history server.

Classificações	Descrição	Ações de reconfiguração
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do 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	Alterar os valores no arquivo log4j.properties do 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	Alterar a configuração do servidor SSL no Hadoop	Not available.
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop	Not available.

Classificações	Descrição	Ações de reconfiguração
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Mude os valores no HBase ambiente.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.	Not available.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.	This classification should not be reconfigured.
hdfs-env	Mude os valores no HDFS ambiente.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classificações	Descrição	Ações de reconfiguração
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Mude os valores no HCatalog ambiente.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Configurações EMR selecionadas pela Amazon para o Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.	Not available.

Classificações	Descrição	Ações de reconfiguração
hive-parquet-logging	Alterar os valores no arquivo <code>parquet-logging.properties</code> do Hive.	Not available.
hive-env	Alterar os valores no ambiente do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Altere os valores no arquivo <code>hive-exec-log 4j2.properties</code> do Hive.	Not available.
hive-llap-daemon-log4j2	Altere os valores no arquivo <code>llap-daemon-log 4j2.properties</code> do Hive.	Not available.
hive-log4j2	Alterar os valores no arquivo <code>hive-log4j2.properties</code> do Hive.	Not available.
hive-site	Alterar os valores no arquivo <code>hive-site.xml</code> do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Alterar os valores no arquivo <code>hiveserver2-site.xml</code> do Hive Server2.	Not available.
hue-ini	Alterar os valores no arquivo <code>ini</code> do Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.

Classificações	Descrição	Ações de reconfiguração
httpfs-env	Mude os valores no HTTPFS ambiente.	Restarts Hadoop Httpfs service.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.	Not available.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.	Not available.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Altere os valores no ambiente do Hudi.	Not available.
hudi-defaults	Alterar os valores no arquivo hudi-defaults.conf do Hudi.	Not available.
iceberg-defaults	Alterar os valores no arquivo iceberg-defaults.conf do Iceberg.	Not available.
delta-defaults	Alterar os valores no arquivo delta-defaults.conf do Delta.	Not available.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.	Not available.

Classificações	Descrição	Ações de reconfiguração
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo <code>jupyterhub_config.py</code> do.	Not available.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.	Not available.
jupyter-sparkmagic-conf	Altere os valores no arquivo <code>config.json</code> do Sparkmagic.	Not available.
livy-conf	Alterar os valores no arquivo <code>livy.conf</code> do Livy.	Restarts Livy Server.
livy-env	Alterar os valores no ambiente do Livy.	Restarts Livy Server.
livy-log4j2	Alterar as configurações de <code>log4j2.properties</code> do Livy.	Restarts Livy Server.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Altere os valores no arquivo <code>mapred-site.xml</code> do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Alterar os valores no ambiente do Oozie.	Restarts Oozie.
oozie-log4j	Alterar os valores no arquivo <code>oozie-log4j.properties</code> do Oozie.	Restarts Oozie.
oozie-site	Alterar os valores no arquivo <code>oozie-site.xml</code> do Oozie.	Restarts Oozie.

Classificações	Descrição	Ações de reconfiguração
phoenix-hbase-metrics	Alterar os valores no arquivo <code>hadoop-metrics2-hbase.properties</code> do Phoenix.	Not available.
phoenix-hbase-site	Alterar os valores no arquivo <code>hbase-site.xml</code> do Phoenix.	Not available.
phoenix-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Alterar os valores no arquivo <code>hadoop-metrics2-phoenix.properties</code> do Phoenix.	Not available.
pig-env	Alterar os valores no ambiente do Pig.	Not available.
pig-properties	Alterar os valores no arquivo <code>pig.properties</code> do Pig.	Restarts Oozie.
pig-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Pig.	Not available.
presto-log	Alterar os valores no arquivo <code>log.properties</code> do Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Alterar os valores no arquivo <code>config.properties</code> do Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Altere os valores no arquivo <code>password-authenticator.properties</code> do Presto.	Not available.
presto-env	Altere os valores no arquivo <code>presto-env.sh</code> do Presto.	Restarts Presto-Server (for PrestoDB)

Classificações	Descrição	Ações de reconfiguração
presto-node	Altere os valores no arquivo node.properties do Presto.	Not available.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.	Not available.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.	Not available.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.	Not available.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.	Not available.
presto-connector-lakeformat ion	Alterar os valores no arquivo lakeformation.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.	Not available.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.	Not available.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.	Not available.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.	Not available.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.	Not available.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.	Not available.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.	Not available.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.	Not available.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.	Not available.
trino-log	Alterar os valores no arquivo log.properties do Trino.	Restarts Trino-Server (for Trino)
trino-config	Alterar os valores no arquivo config.properties do Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Alterar os valores no arquivo password-authenticator.properties do Trino.	Restarts Trino-Server (for Trino)
trino-env	Alterar os valores no arquivo trino-env.sh do Trino.	Restarts Trino-Server (for Trino)
trino-node	Alterar os valores no arquivo node.properties do Trino.	Not available.

Classificações	Descrição	Ações de reconfiguração
trino-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Trino.	Not available.
trino-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Trino.	Not available.
trino-connector-delta	Alterar os valores no arquivo delta.properties do Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Alterar os valores no arquivo hive.properties do Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Alterar os valores no arquivo exchange-manager.properties do Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Alterar os valores no arquivo iceberg.properties do Trino.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Alterar os valores no arquivo hudi.properties do Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Alterar os valores no arquivo jmx.properties do Trino.	Not available.
trino-connector-kafka	Alterar os valores no arquivo kafka.properties do Trino.	Not available.
trino-connector-localfile	Alterar os valores no arquivo localfile.properties do Trino.	Not available.
trino-connector-memory	Alterar os valores no arquivo memory.properties do Trino.	Not available.
trino-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Trino.	Not available.

Classificações	Descrição	Ações de reconfiguração
trino-connector-mysql	Alterar os valores no arquivo mysql.properties do Trino.	Not available.
trino-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Trino.	Not available.
trino-connector-raptor	Alterar os valores no arquivo raptor.properties do Trino.	Not available.
trino-connector-redis	Alterar os valores no arquivo redis.properties do Trino.	Not available.
trino-connector-redshift	Alterar os valores no arquivo redshift.properties do Trino.	Not available.
trino-connector-tpch	Alterar os valores no arquivo tpch.properties do Trino.	Not available.
trino-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Trino.	Not available.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.	Restarts Ranger KMS Server.
ranger-kms-logback	Altere os valores no arquivo kms-logback.xml do Ranger. KMS	Not available.

Classificações	Descrição	Ações de reconfiguração
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.	Not available.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Alterar os valores no ambiente do Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.	Not available.
spark-log4j2	Alterar os valores no arquivo log4j2.properties do Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Alterar os valores no ambiente do Sqoop.	Not available.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.	Not available.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.	Not available.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.	Restart Oozie and HiveServer2.

Classificações	Descrição	Ações de reconfiguração
yarn-env	Mude os valores no YARN ambiente.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.	Restarts Zeppelin.
zeppelin-site	Alterar as definições da configuração em zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.	Restarts Zookeeper server.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.	Restarts Zookeeper server.

Log de alterações 6.11.1

Log de alterações da versão 6.11.1 e das notas da versão

Data	Evento	Descrição
2023-08-30	Atualizar notas da versão	Adicionadas várias correções relacionadas ao ambiente de gerenciamento às notas da versão
2023-08-21	Publicação de documentos	Notas de lançamento do Amazon EMR 6.11.1 publicadas pela primeira vez
2023-08-16	Implantação concluída	Amazon EMR 6.11.1 totalmente implantado em todas as regiões suportadas
2023-08-04	Lançamento inicial	Amazon EMR 6.11.1 foi implantado pela primeira vez em regiões comerciais limitadas

Amazon EMR versão 6.11.0

Versões da aplicação 6.11.0

As seguintes aplicações são compatíveis com esta versão: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)

- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-6.11.0	emr-6.10.1	emr-6.10.0	emr-6.9.1
AWS SDK para 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	2.12.15	2.12.15	2.12.15	2.12.15
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
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

	emr-6.11.0	emr-6.10.1	emr-6.10.0	emr-6.9.1
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	0.10.2	0.10.2	0.10.2	0.10.2
Trino (PrestoSQL)	410	403	403	398
Zeppelin	0.10.1	0.10.1	0.10.1	0.10.1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

Notas da versão 6.11.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 6.11.0 da Amazon. As alterações são referentes à versão 6.10.0. Para obter informações sobre o cronograma da versão, consulte o [log de alterações](#).

Novos atributos

- O Amazon EMR 6.11.0 é compatível com 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 1.2.0-amzn-0, Trino 410-amzn-0 e PrestoDB 0.279-amzn-0. CUDA

Alterações, melhorias e problemas resolvidos

- Com o Amazon EMR 6.11.0, o conector do DynamoDB foi atualizado para a versão 5.0.0. A versão 5.0.0 usa AWS SDK for Java 2.x. As versões anteriores usavam AWS SDK for Java 1.x. Como resultado dessa atualização, recomendamos que você teste seu código antes de usar o conector do DynamoDB com o Amazon 6.11. EMR
- Quando o conector do DynamoDB para EMR Amazon 6.11.0 chama o serviço do DynamoDB, ele usa o valor da região que você fornece para a propriedade `dynamodb.endpoint`. Recomendamos que você também configure `dynamodb.region` quando usar `dynamodb.endpoint` e que ambas as propriedades tenham como destino a mesma Região da AWS. Se você usar `dynamodb.endpoint` e não configurar `dynamodb.region`, o conector do DynamoDB para EMR Amazon 6.11.0 retornará uma exceção de região inválida e tentará reconciliar suas Região da AWS informações do serviço de metadados de instância da Amazon (EC2 IMDS). Se o conector não conseguir recuperar a região IMDS, o padrão é Leste dos EUA (Norte da Virgínia) (`us-east-1`). O erro a seguir é um exemplo da exceção de região inválida que você pode obter se não configurar adequadamente a `dynamodb.region` propriedade:
`error software.amazon.awssdk.services.dynamodb.model.DynamoDbException: Credential should be scoped to a valid region.` Para obter mais informações sobre as classes afetadas pela AWS SDK for Java atualização para 2.x, consulte o commit [Upgrade AWS SDK for Java from 1.x to 2.x \(#175\)](#) no GitHub repositório do conector Amazon EMR - DynamoDB.
- Esta versão corrige um problema em que os dados da coluna se tornam NULL quando você usa o Delta Lake para armazenar dados da tabela Delta no Amazon S3 após a operação de renomeação da coluna. Para obter mais informações sobre esse atributo experimental no Delta Lake, consulte [Operação de renomeação de coluna](#) no Guia do usuário do Delta Lake.
- A versão 6.11.0 corrige um problema que pode ocorrer quando você cria um nó de borda ao replicar um dos nós primários de um cluster com vários nós primários. O nó de borda replicado pode causar atrasos nas operações de redução da escala verticalmente ou resultar em alta utilização de memória nos nós primários. Para obter mais informações sobre como criar um nó de

borda para se comunicar com seu EMR cluster, consulte [Edge Node Creator](#) no `aws-samples` repositório em GitHub.

- A versão 6.11.0 melhora o processo de automação que a Amazon EMR usa para remontar volumes da EBS Amazon em uma instância após uma reinicialização.
- A versão 6.11.0 corrige um problema que resultou em lacunas intermitentes nas métricas do Hadoop que a Amazon publica na Amazon. EMR CloudWatch
- A versão 6.11.0 corrige um problema com EMR clusters em que uma atualização no arquivo de YARN configuração que contém a lista de exclusão de nós do cluster é interrompida devido à utilização excessiva do disco. A atualização incompleta impede futuras operações de redução da escala verticalmente do cluster. Esta versão garante que o cluster permaneça íntegro e que as operações de escalabilidade funcionem conforme esperado.
- O tamanho padrão do volume raiz aumentou para 15 GB no Amazon EMR 6.10.0 e superior. O tamanho padrão do volume raiz das versões anteriores é de 10 GB.
- O Hadoop 3.3.3 introduziu uma alteração em YARN ([YARN-9608](#)) que mantém os nós onde os contêineres eram executados em um estado de descomissionamento até que o aplicativo seja concluído. Essa alteração garante que dados locais, como dados embaralhados, não sejam perdidos e que você não precise executar o trabalho novamente. Essa abordagem também pode levar à subutilização de recursos em clusters com ou sem o ajuste de escala gerenciado habilitado.

Com as EMR versões 6.11.0 e superiores da Amazon, bem como 6.8.1, 6.9.1 e 6.10.1, o valor de `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` é definido como `in` para resolver esse problema. `false` `yarn-site.xml`

Embora a correção resolva os problemas introduzidos pelo YARN -9608, ela pode fazer com que as tarefas do Hive falhem devido à perda aleatória de dados em clusters com escalabilidade gerenciada habilitada. Reduzimos esse risco nesta versão também ao configurar `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-shuffle-data` para workloads do Hive. Essa configuração só está disponível nas EMR versões 6.11.0 e superiores da Amazon.

- Quando você inicia um cluster com a versão de patch mais recente do Amazon EMR 5.36 ou superior, 6.6 ou superior ou 7.0 ou superior, a Amazon EMR usa a versão mais recente do Amazon Linux 2023 ou Amazon Linux 2 para a Amazon padrão. EMR AMI Para obter mais informações, consulte [Usando o Amazon Linux padrão AMI para a Amazon EMR](#).

Note

Esta versão não recebe mais AMI atualizações automáticas, pois foi sucedida por mais 1 lançamento de patch. A versão de patch é indicada pelo número após o segundo ponto decimal (6.8.**1**). Para ver se você está usando a versão de patch mais recente, verifique as versões disponíveis no [Guia de lançamento](#) ou verifique o menu suspenso de EMRlançamentos da Amazon ao criar um cluster no console ou usar a ação [ListReleaseLabels](#)API ou [list-release-labels](#)CLI. Para receber atualizações sobre novos lançamentos, assine o RSS feed do [What's new?](#) página.

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 709.1	4.14.348	23 de julho de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia), Ásia

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Pacífico (Hyderabad), Oriente Médio (), Europa (EspanhaUAE), Europa (Zurique), Ásia-Pacífico (Melbourne), Israel (Tel Aviv), Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 808.0	4.14.320	24 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), Israel (Tel Aviv).

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 727.0	4.14.320	14 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 719.0	4.14.320	2 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)
2.0.2023 628.0	4.14.318	12 de julho de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Zurique), Europa (Milão), Europa (Espanha), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad)), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio () UAE

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 612.0	4.14.314	23 de junho de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Zurique), Europa (Milão), Europa (Espanha), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad)), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio () UAE

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 504.1	4.14.313	16 de maio de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Oriente Médio (), Canadá (Central) UAE

Versões de componente 6.11.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.4.2	Amazon SageMaker Spark SDK
<code>delta</code>	2.2.0	O Delta Lake é um formato de tabela aberto para grandes conjuntos de dados analíticos
<code>delta-standalone-connectors</code>	0.6.0	Os conectores do Delta oferecem runtimes diferentes para integrar o Delta Lake a mecanismos, como Flink, Hive e Presto.
<code>emr-ddb</code>	5.1.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	3.4.0	Bibliotecas convenientes para o ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
emr-kinesis	3.8.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-notebook-env	1.7.0	Ambiente Conda para bloco de anotações do emr, que inclui o jupyter enterprise gateway
emr-s3-dist-cp	2.25.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	2.4.0	EMRConector S3Select
emr-wal-cli	1.1.0	Cli usada para lista/exclusão de emrwal.
emrfs	2.56.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.16.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
flink-jobmanager-config	1.16.0	Gerenciando recursos em EMR nós para o Apache JobManager Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.

Componente	Version (Versão)	Descrição
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	3.3.3-amzn-3	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	3.3.3-amzn-3	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	3.3.3-amzn-3	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	3.3.3-amzn-3	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	3.3.3-amzn-3	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httpfs-server	3.3.3-amzn-3	HTTPendpoint para HDFS operações.
hadoop-kms-server	3.3.3-amzn-3	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API

Componente	Version (Versão)	Descrição
hadoop-mapred	3.3.3-amzn-3	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	3.3.3-amzn-3	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	3.3.3-amzn-3	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	3.3.3-amzn-3	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	2.4.15-amzn-1	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	2.4.15-amzn-1	Serviço para atender a uma ou mais HBase regiões.
hbase-client	2.4.15-amzn-1	HBasecliente de linha de comando.
hbase-rest-server	2.4.15-amzn-1	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	2.4.15-amzn-1	Serviço que fornece um endpoint Thrift para. HBase

Componente	Version (Versão)	Descrição
hbase-operator-tools	2.4.15-amzn-1	Ferramenta de reparo para HBase clusters Apache.
hcatalog-client	3.1.3-amzn-4	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	3.1.3-amzn-4	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	3.1.3-amzn-4	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	3.1.3-amzn-4	O cliente da linha de comando do Hive.
hive-hbase	3.1.3-amzn-4	Cliente do Hive-hbase.
hive-metastore-server	3.1.3-amzn-4	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	3.1.3-amzn-4	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.13.0-amzn-0	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.

Componente	Version (Versão)	Descrição
hudi-presto	0.13.0-amzn-0	Biblioteca de pacotes para executar o Presto com o Hudi.
hudi-trino	0.13.0-amzn-0	Biblioteca de pacotes para execução do Presto com o Hudi.
hudi-spark	0.13.0-amzn-0	Biblioteca de pacotes para executar o Spark com o Hudi.
hue-server	4.11.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
iceberg	1.2.0-amzn-0	Apache Iceberg é um formato de tabela aberta para conjuntos de dados analíticos imensos
jupyterhub	1.4.1	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.7.1-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mxnet	1.9.1	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mariadb-server	5.5.68+	Servidor de banco de dados MariaDB.
nvidia-cuda	11.8.0	Drivers NVIDIA e toolkit CUDA

Componente	Version (Versão)	Descrição
oozie-client	5.2.1	O cliente da linha de comando do Oozie.
oozie-server	5.2.1	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	4.5.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	5.1.2	As bibliotecas do phoenix para servidor e cliente
phoenix-connectors	5.1.2	Conectores do Apache Phoenix para Spark-3
phoenix-query-server	5.1.2	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.279-amzn-0	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.279-amzn-0	O serviço que executa partes de uma consulta.
presto-client	0.279-amzn-0	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.

Componente	Version (Versão)	Descrição
trino-coordinator	410-amzn-0	Serviço que aceita consultas e gerencia a execução de consultas entre trino-workers.
trino-worker	410-amzn-0	O serviço que executa partes de uma consulta.
trino-client	410-amzn-0	Cliente de linha de comando do Trino instalado em mestres de espera de um cluster de alta disponibilidade (HA) em que o servidor do Trino não é iniciado.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	4.0.2	O projeto R para computação estatística
ranger-kms-server	2.0.0	Sistema de gerenciamento de chaves Apache Ranger
spark-client	3.3.2-amzn-0	Os clientes da linha de comando do Spark.
spark-history-server	3.3.2-amzn-0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	3.3.2-amzn-0	Mecanismo de execução na memória para YARN.

Componente	Version (Versão)	Descrição
spark-yarn-slave	3.3.2-amzn-0	Bibliotecas do Apache Spark necessárias para escravos. YARN
spark-rapids	23.02.0-amzn-0	RAPIDSPlugin Nvidia Spark que acelera o Apache Spark com o. GPUs
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	2.11.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.10.2-amzn-2	O YARN aplicativo e as bibliotecas tez.
tez-on-worker	0.10.2-amzn-2	O YARN aplicativo tez e as bibliotecas para nós de trabalho.
webserver	2.4.41+	HTTPServidor Apache.
zeppelin-server	0.10.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.5.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.

Componente	Version (Versão)	Descrição
zookeeper-client	3.5.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 6.11.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Ocorrem ações de reconfiguração quando você especifica uma configuração para grupos de instâncias em um cluster em execução. A Amazon EMR só inicia ações de reconfiguração para as classificações que você modifica. Para ter mais informações, consulte [Reconfigurar um grupo de instâncias de um cluster em execução](#).

Classificações do emr-6.11.0

Classificações	Descrição	Ações de reconfiguração
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.	Restarts the ResourceManager service.
container-executor	Altere os valores no arquivo <code>container-executor.cfg</code> YARN do Hadoop.	Not available.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.	Not available.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager,

Classificações	Descrição	Ações de reconfiguração
		NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Alterar as configurações relacionadas ao docker.	Not available.
emrfs-site	Altere EMRFS as configurações.	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	Alterar as configurações do flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Alterar as configurações de log4j.properties no Flink.	Restarts Flink history server.

Classificações	Descrição	Ações de reconfiguração
flink-log4j-session	Alterar as configurações de log4j-yarn-session.properties no Flink para sessão do Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.	Restarts Flink history server.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do 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	Alterar os valores no arquivo log4j.properties do 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.

Classificações	Descrição	Ações de reconfiguração
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop	Not available.
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop	Not available.
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Mude os valores no HBase ambiente.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.	Not available.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.	This classification should not be reconfigured.

Classificações	Descrição	Ações de reconfiguração
hdfs-env	Mude os valores no HDFS ambiente.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Mude os valores no HCatalog ambiente.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Configurações EMR selecionadas pela Amazon para o Apache Hive.	Sets configurations to launch Hive LLAP service.

Classificações	Descrição	Ações de reconfiguração
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.	Not available.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.	Not available.
hive-env	Alterar os valores no ambiente do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log 4j2.properties do Hive.	Not available.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log 4j2.properties do Hive.	Not available.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.	Not available.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.	Not available.

Classificações	Descrição	Ações de reconfiguração
hue-ini	Alterar os valores no arquivo ini do Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Mude os valores no HTTPFS ambiente.	Restarts Hadoop Httpfs service.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.	Not available.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.	Not available.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Altere os valores no ambiente do Hudi.	Not available.
hudi-defaults	Alterar os valores no arquivo hudi-defaults.conf do Hudi.	Not available.
iceberg-defaults	Alterar os valores no arquivo iceberg-defaults.conf do Iceberg.	Not available.
delta-defaults	Alterar os valores no arquivo delta-defaults.conf do Delta.	Not available.

Classificações	Descrição	Ações de reconfiguração
jupyter-notebook-conf	Alterar os valores no arquivo <code>jupyter_notebook_config.py</code> do Notebook Jupyter.	Not available.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo <code>jupyterhub_config.py</code> do.	Not available.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.	Not available.
jupyter-sparkmagic-conf	Altere os valores no arquivo <code>config.json</code> do Sparkmagic.	Not available.
livy-conf	Alterar os valores no arquivo <code>livy.conf</code> do Livy.	Restarts Livy Server.
livy-env	Alterar os valores no ambiente do Livy.	Restarts Livy Server.
livy-log4j2	Alterar as configurações de <code>log4j2.properties</code> do Livy.	Restarts Livy Server.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Altere os valores no arquivo <code>mapred-site.xml</code> do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Alterar os valores no ambiente do Oozie.	Restarts Oozie.
oozie-log4j	Alterar os valores no arquivo <code>oozie-log4j.properties</code> do Oozie.	Restarts Oozie.

Classificações	Descrição	Ações de reconfiguração
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.	Not available.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.	Not available.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.	Not available.
pig-env	Alterar os valores no ambiente do Pig.	Not available.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.	Restarts Oozie.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.	Not available.
presto-log	Alterar os valores no arquivo log.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Alterar os valores no arquivo config.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-env	Altere os valores no arquivo presto-env.sh do Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Altere os valores no arquivo node.properties do Presto.	Not available.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.	Not available.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.	Not available.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.	Not available.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.	Not available.
presto-connector-lakeformat ion	Alterar os valores no arquivo lakeformation.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.	Not available.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.	Not available.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.	Not available.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.	Not available.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.	Not available.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.	Not available.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.	Not available.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.	Not available.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.	Not available.
trino-log	Alterar os valores no arquivo log.properties do Trino.	Restarts Trino-Server (for Trino)
trino-config	Alterar os valores no arquivo config.properties do Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Alterar os valores no arquivo password-authenticator.properties do Trino.	Restarts Trino-Server (for Trino)
trino-env	Alterar os valores no arquivo trino-env.sh do Trino.	Restarts Trino-Server (for Trino)

Classificações	Descrição	Ações de reconfiguração
trino-node	Alterar os valores no arquivo <code>node.properties</code> do Trino.	Not available.
trino-connector-blackhole	Alterar os valores no arquivo <code>blackhole.properties</code> do Trino.	Not available.
trino-connector-cassandra	Alterar os valores no arquivo <code>cassandra.properties</code> do Trino.	Not available.
trino-connector-delta	Alterar os valores no arquivo <code>delta.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Alterar os valores no arquivo <code>hive.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Alterar os valores no arquivo <code>exchange-manager.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Alterar os valores no arquivo <code>iceberg.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Alterar os valores no arquivo <code>hudi.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Alterar os valores no arquivo <code>jmx.properties</code> do Trino.	Not available.
trino-connector-kafka	Alterar os valores no arquivo <code>kafka.properties</code> do Trino.	Not available.
trino-connector-localfile	Alterar os valores no arquivo <code>localfile.properties</code> do Trino.	Not available.
trino-connector-memory	Alterar os valores no arquivo <code>memory.properties</code> do Trino.	Not available.

Classificações	Descrição	Ações de reconfiguração
trino-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Trino.	Not available.
trino-connector-mysql	Alterar os valores no arquivo mysql.properties do Trino.	Not available.
trino-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Trino.	Not available.
trino-connector-raptor	Alterar os valores no arquivo raptor.properties do Trino.	Not available.
trino-connector-redis	Alterar os valores no arquivo redis.properties do Trino.	Not available.
trino-connector-redshift	Alterar os valores no arquivo redshift.properties do Trino.	Not available.
trino-connector-tpch	Alterar os valores no arquivo tpch.properties do Trino.	Not available.
trino-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Trino.	Not available.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.	Restarts Ranger KMS Server.
ranger-kms-logback	Altere os valores no arquivo kms-logback.xml do Ranger. KMS	Not available.

Classificações	Descrição	Ações de reconfiguração
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.	Not available.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Alterar os valores no ambiente do Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.	Not available.
spark-log4j2	Alterar os valores no arquivo log4j2.properties do Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Alterar os valores no ambiente do Sqoop.	Not available.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.	Not available.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.	Not available.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.	Restart Oozie and HiveServer2.

Classificações	Descrição	Ações de reconfiguração
yarn-env	Mude os valores no YARN ambiente.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.	Restarts Zeppelin.
zeppelin-site	Alterar as definições da configuração em zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.	Restarts Zookeeper server.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.	Restarts Zookeeper server.

Log de alterações 6.11.0

Log de alterações da versão 6.11.0 e das notas da versão

Data	Evento	Descrição
2023-08-21	Atualizar	Problema corrigido introduzido com o Hadoop 3.3.3.
2023-07-26	Atualizar	Novos rótulos de versão 2.0.20230612.0 e 2.0.20230628.0 do sistema operacional.
2023-06-09	Implantação concluída	Amazon EMR 6.11.0 totalmente implantado em todas as regiões suportadas
2023-06-09	Publicação de documentos	Notas de lançamento do Amazon EMR 6.11.0 publicadas pela primeira vez
2023-06-08	Lançamento inicial	Amazon EMR 6.11.0 foi implantado pela primeira vez nas regiões comerciais iniciais

Amazon EMR versão 6.10.1

Versões 6.10.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-6.10.1	emr-6.10.0	emr-6.9.1	emr-6.9.0
AWS SDK para 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	2.12.15	2.12.15	2.12.15	2.12.15
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

	emr-6.10.1	emr-6.10.0	emr-6.9.1	emr-6.9.0
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
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	0.10.2	0.10.2	0.10.2	0.10.2
Trino (PrestoSQL)	403	403	398	398
Zeppelin	0.10.1	0.10.1	0.10.1	0.10.1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

Notas da versão 6.10.1

As notas de lançamento a seguir incluem informações sobre a EMR versão 6.10.1 da Amazon. As alterações são referentes à versão 6.10.0. Para obter informações sobre o cronograma da versão, consulte o [Log de alterações 6.10.1](#).

Alterações, melhorias e problemas resolvidos

- Devido à contenção de bloqueio, um nó pode entrar em um deadlock se for adicionado ou removido ao mesmo tempo em que tenta ser desativado. Como resultado, o Hadoop Resource Manager (YARN) deixa de responder e afeta todos os contêineres recebidos e em execução no momento.
- O Hadoop 3.3.3 introduziu uma alteração em YARN ([YARN-9608](#)) que mantém os nós onde os contêineres eram executados em um estado de descomissionamento até que o aplicativo seja concluído. Essa alteração garante que dados locais, como dados embaralhados, não sejam perdidos e que você não precise executar o trabalho novamente. Essa abordagem também pode levar à subutilização de recursos em clusters com ou sem o ajuste de escala gerenciado habilitado.

Com as EMR versões 6.11.0 e superiores da Amazon, bem como 6.8.1, 6.9.1 e 6.10.1, o valor de `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` é definido como `true` para resolver esse problema. `false` `yarn-site.xml`

Embora a correção resolva os problemas introduzidos pelo YARN -9608, ela pode fazer com que as tarefas do Hive falhem devido à perda aleatória de dados em clusters com escalabilidade gerenciada habilitada. Reduzimos esse risco nesta versão também ao configurar `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-shuffle-data` para workloads do Hive. Essa configuração só está disponível nas EMR versões 6.11.0 e superiores da Amazon.

- O coletor de métricas não enviará métricas para o ambiente de gerenciamento após o failover do nó primário em clusters com a configuração de grupos de instâncias.
- Esta versão inclui uma alteração que permite que clusters de alta disponibilidade se recuperem de um estado de falha após a reinicialização.
- Esta versão inclui correções de segurança para Hue e HBase
- Esta versão corrige um problema em que clusters que estão executando cargas de trabalho no Spark com a Amazon EMR podem receber silenciosamente resultados incorretos com `contains`, `startsWith`, e `endsWith` `like` Esse problema ocorre quando você usa as expressões em campos particionados que têm metadados no Amazon EMR Hive3 Metastore Server (`hms`).
- Esta versão corrige um problema de limitação no lado do Glue quando não há funções definidas pelo usuário (`udf`).

- Esta versão corrige um problema que exclui registros de contêineres pelo serviço de agregação de registros de nós antes que o carregador de registros possa enviá-los para o S3 em caso de descomissionamento. YARN
- Esta versão corrige um problema com as métricas do FairShare Scheduler quando o Node Label está habilitado para o Hadoop.
- Esta versão corrige um problema que afetou a performance do Spark quando você definiu um valor de `true` padrão para a configuração `spark.yarn.heterogeneousExecutors.enabled` no `spark-defaults.conf`.
- Esta versão corrige um problema com a falha do Reduce Task em ler dados embaralhados. O problema causou falhas na consulta do Hive com um erro de memória corrompida.
- Esta versão adiciona um novo mecanismo de repetição ao fluxo de trabalho de escalabilidade de EMR clusters para clusters que executam o Presto ou o Trino. Essa melhoria reduz o risco de que o redimensionamento do cluster fique paralisado indefinidamente devido a uma única falha na operação de redimensionamento. Ela também aprimora a utilização dos clusters, porque seu cluster aumenta e reduz a escala verticalmente com mais rapidez.
- Esta versão melhora a lógica de redução do cluster para que seu cluster não tente reduzir os nós principais abaixo da configuração do fator de HDFS replicação do cluster. Isso se alinha aos seus requisitos de redundância de dados e reduz a probabilidade de uma operação de escalabilidade paralisar.
- O daemon de gerenciamento de logs foi atualizado para identificar todos os logs que estão em uso ativo com identificadores de arquivos abertos no armazenamento da instância local e nos processos associados. Essa atualização garante que a Amazon exclua EMR adequadamente os arquivos e recupere o espaço de armazenamento depois que os registros forem arquivados no Amazon S3.
- Esta versão inclui um aprimoramento do daemon de gerenciamento de logs que exclui diretórios de etapas vazios e não utilizados no sistema de arquivos de cluster local. Um número excessivamente grande de diretórios vazios pode degradar o desempenho dos EMR daemons da Amazon e resultar na sobreutilização do disco.
- Esta versão corrige um problema que pode ocorrer quando você cria um nó de borda ao replicar um dos nós primários de um cluster com vários nós primários. O nó de borda replicado pode causar atrasos nas operações de redução da escala verticalmente ou resultar em alta utilização de memória nos nós primários. Para obter mais informações sobre como criar um nó de borda para se comunicar com seu EMR cluster, consulte [Edge Node Creator](#) no `aws-samples` repositório em GitHub.

- Esta versão melhora o processo de automação que a Amazon EMR usa para remontar EBS volumes da Amazon em uma instância após uma reinicialização.
- Esta versão corrige um problema que resultou em lacunas intermitentes nas métricas do Hadoop que a Amazon publica EMR na Amazon. CloudWatch
- Esta versão corrige um problema com EMR clusters em que uma atualização no arquivo de YARN configuração que contém a lista de exclusão de nós do cluster é interrompida devido à utilização excessiva do disco. A atualização incompleta impede futuras operações de redução da escala verticalmente do cluster. Esta versão garante que o cluster permaneça íntegro e que as operações de escalabilidade funcionem conforme esperado.
- Quando você inicia um cluster com a versão de patch mais recente do Amazon EMR 5.36 ou superior, 6.6 ou superior ou 7.0 ou superior, a Amazon EMR usa a versão mais recente do Amazon Linux 2023 ou Amazon Linux 2 para a Amazon padrão. EMR AMI Para obter mais informações, consulte [Usando o Amazon Linux padrão AMI para a Amazon EMR](#).

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 709.1	4.14.348	23 de julho de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			<p>Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia), Ásia Pacífico (Hyderabad), Oriente Médio (), Europa (EspanhaUAE), Europa (Zurique), Ásia-Pacífico (Melbourne), Israel (Tel Aviv), Oeste do Canadá (Calgary)</p>

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 223.0	4.14.336	8 de março de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelease Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia) , Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 131.0	4.14.336	14 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelease Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia) , Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 124.0	4.14.336	7 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia) , Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 109.0	4.14.334	24 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia) , Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 218.0	4.14.330	2 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 206.0	4.14.330	22 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelease Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 116.0	4.14.328	11 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 101.0	4.14.327	16 de novembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 020.1	4.14.326	7 de novembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 012.1	4.14.326	26 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelease Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 926.0	4.14.322	19 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 906.0	4.14.322	4 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 822.0	4.14.322	30 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 808.0	4.14.320	24 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)
2.0.2023 727.0	4.14.320	14 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central)

Versões de componente 6.10.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR

e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-`amzn`-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.4.2	Amazon SageMaker Spark SDK
<code>delta</code>	2.2.0	O Delta Lake é um formato de tabela aberto para grandes conjuntos de dados analíticos
<code>emr-ddb</code>	4.16.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	3.3.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.7.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-notebook-env</code>	1.7.0	Ambiente Conda para bloco de anotações do emr, que inclui o jupyter enterprise gateway
<code>emr-s3-dist-cp</code>	2.24.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.

Componente	Version (Versão)	Descrição
emr-s3-select	2.3.0	EMRConector S3Select
emr-wal-cli	1.0.0	Cli usada para lista/exclusão de emrwal.
emrfs	2.55.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.16.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
flink-jobmanager-config	1.16.0	Gerenciando recursos em EMR nós para o Apache JobManager Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	3.3.3-amzn-2.1	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.

Componente	Version (Versão)	Descrição
hadoop-hdfs-datanode	3.3.3-amzn-2.1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	3.3.3-amzn-2.1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	3.3.3-amzn-2.1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	3.3.3-amzn-2.1	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-https-server	3.3.3-amzn-2.1	HTTPendpoint para HDFS operações.
hadoop-kms-server	3.3.3-amzn-2.1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	3.3.3-amzn-2.1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	3.3.3-amzn-2.1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	3.3.3-amzn-2.1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	3.3.3-amzn-2.1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.

Componente	Version (Versão)	Descrição
hbase-hmaster	2.4.15-amzn-0.1	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	2.4.15-amzn-0.1	Serviço para atender a uma ou mais HBase regiões.
hbase-client	2.4.15-amzn-0.1	HBase cliente de linha de comando.
hbase-rest-server	2.4.15-amzn-0.1	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	2.4.15-amzn-0.1	Serviço que fornece um endpoint Thrift para HBase.
hbase-operator-tools	2.4.15-amzn-0.1	Ferramenta de reparo para HBase clusters Apache.
hcatalog-client	3.1.3-amzn-3.1	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	3.1.3-amzn-3.1	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	3.1.3-amzn-3.1	HTTP endpoint fornecendo uma REST interface para HCatalog.

Componente	Version (Versão)	Descrição
hive-client	3.1.3-amzn-3.1	O cliente da linha de comando do Hive.
hive-hbase	3.1.3-amzn-3.1	Cliente do Hive-hbase.
hive-metastore-server	3.1.3-amzn-3.1	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	3.1.3-amzn-3.1	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.12.2-amzn-0	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.
hudi-presto	0.12.2-amzn-0	Biblioteca de pacotes para executar o Presto com o Hudi.
hudi-trino	0.12.2-amzn-0	Biblioteca de pacotes para execução do Presto com o Hudi.
hudi-spark	0.12.2-amzn-0	Biblioteca de pacotes para executar o Spark com o Hudi.
hue-server	4.10.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop

Componente	Version (Versão)	Descrição
iceberg	1.1.0-amzn-0	Apache Iceberg é um formato de tabela aberta para conjuntos de dados analíticos imensos
jupyterhub	1.5.0	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.7.1-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mxnet	1.9.1	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mariadb-server	5.5.68+	Servidor de banco de dados MariaDB.
nvidia-cuda	11.8.0	Drivers NVIDIA e toolkit CUDA
oozie-client	5.2.1	O cliente da linha de comando do Oozie.
oozie-server	5.2.1	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	4.5.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	5.1.2	As bibliotecas do phoenix para servidor e cliente

Componente	Version (Versão)	Descrição
phoenix-connectors	6.0.0- SNAPSHOT	Conectores do Apache Phoenix para Spark-3
phoenix-query-server	6.0.0	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.278.1-amzn-0	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.278.1-amzn-0	O serviço que executa partes de uma consulta.
presto-client	0.278.1-amzn-0	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
trino-coordinator	403-amzn-0	Serviço que aceita consultas e gerencia a execução de consultas entre trino-workers.
trino-worker	403-amzn-0	O serviço que executa partes de uma consulta.
trino-client	403-amzn-0	Cliente de linha de comando do Trino instalado em mestres de espera de um cluster de alta disponibilidade (HA) em que o servidor do Trino não é iniciado.

Componente	Version (Versão)	Descrição
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	4.0.2	O projeto R para computação estatística
ranger-kms-server	2.0.0	Sistema de gerenciamento de chaves Apache Ranger
spark-client	3.3.1-amzn-0.1	Os clientes da linha de comando do Spark.
spark-history-server	3.3.1-amzn-0.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	3.3.1-amzn-0.1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	3.3.1-amzn-0.1	Bibliotecas do Apache Spark necessárias para escravos. YARN
spark-rapids	22.12.0-amzn-0	RAPIDSPlugin Nvidia Spark que acelera o Apache Spark com o. GPUs
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	2.11.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.

Componente	Version (Versão)	Descrição
tez-on-yarn	0.10.2-amzn-1.1	O YARN aplicativo e as bibliotecas tez.
tez-on-worker	0.10.2-amzn-1.1	O YARN aplicativo tez e as bibliotecas para nós de trabalho.
webserver	2.4.41+	HTTPServidor Apache.
zeppelin-server	0.10.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.5.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.5.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 6.10.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Ocorrem ações de reconfiguração quando você especifica uma configuração para grupos de instâncias em um cluster em execução. A Amazon EMR só inicia ações de reconfiguração para as classificações que você modifica. Para ter mais informações, consulte [Reconfigurar um grupo de instâncias de um cluster em execução](#).

Classificações do emr-6.10.1

Classificações	Descrição	Ações de reconfiguração
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.	Restarts the Resource Manager service.
container-executor	Altere os valores no arquivo container-executor.cfg YARN do Hadoop.	Not available.
container-log4j	Altere os valores no arquivo container-log4j.properties YARN do Hadoop.	Not available.
core-site	Alterar os valores no arquivo core-site.xml do 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	Alterar as configurações relacionadas ao docker.	Not available.
emrfs-site	Altere EMRFS as configurações.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

Classificações	Descrição	Ações de reconfiguração
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Alterar as configurações do flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Alterar as configurações de log4j.properties no Flink.	Restarts Flink history server.
flink-log4j-session	Alterar as configurações de log4j-yarn-session.properties no Flink para sessão do Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.	Restarts Flink history server.

Classificações	Descrição	Ações de reconfiguração
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do 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	Alterar os valores no arquivo log4j.properties do 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	Alterar a configuração do servidor SSL no Hadoop	Not available.
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop	Not available.

Classificações	Descrição	Ações de reconfiguração
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Mude os valores no HBase ambiente.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.	Not available.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.	This classification should not be reconfigured.
hdfs-env	Mude os valores no HDFS ambiente.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classificações	Descrição	Ações de reconfiguração
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Mude os valores no HCatalog ambiente.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Configurações EMR selecionadas pela Amazon para o Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.	Not available.

Classificações	Descrição	Ações de reconfiguração
hive-parquet-logging	Alterar os valores no arquivo <code>parquet-logging.properties</code> do Hive.	Not available.
hive-env	Alterar os valores no ambiente do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Altere os valores no arquivo <code>hive-exec-log 4j2.properties</code> do Hive.	Not available.
hive-llap-daemon-log4j2	Altere os valores no arquivo <code>llap-daemon-log 4j2.properties</code> do Hive.	Not available.
hive-log4j2	Alterar os valores no arquivo <code>hive-log4j2.properties</code> do Hive.	Not available.
hive-site	Alterar os valores no arquivo <code>hive-site.xml</code> do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Alterar os valores no arquivo <code>hiveserver2-site.xml</code> do Hive Server2.	Not available.
hue-ini	Alterar os valores no arquivo <code>ini</code> do Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.

Classificações	Descrição	Ações de reconfiguração
httpfs-env	Mude os valores no HTTPFS ambiente.	Restarts Hadoop Httpfs service.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.	Not available.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.	Not available.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Altere os valores no ambiente do Hudi.	Not available.
hudi-defaults	Alterar os valores no arquivo hudi-defaults.conf do Hudi.	Not available.
iceberg-defaults	Alterar os valores no arquivo iceberg-defaults.conf do Iceberg.	Not available.
delta-defaults	Alterar os valores no arquivo delta-defaults.conf do Delta.	Not available.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.	Not available.

Classificações	Descrição	Ações de reconfiguração
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo <code>jupyterhub_config.py</code> do.	Not available.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.	Not available.
jupyter-sparkmagic-conf	Altere os valores no arquivo <code>config.json</code> do Sparkmagic.	Not available.
livy-conf	Alterar os valores no arquivo <code>livy.conf</code> do Livy.	Restarts Livy Server.
livy-env	Alterar os valores no ambiente do Livy.	Restarts Livy Server.
livy-log4j2	Alterar as configurações de <code>log4j2.properties</code> do Livy.	Restarts Livy Server.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Altere os valores no arquivo <code>mapred-site.xml</code> do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Alterar os valores no ambiente do Oozie.	Restarts Oozie.
oozie-log4j	Alterar os valores no arquivo <code>oozie-log4j.properties</code> do Oozie.	Restarts Oozie.
oozie-site	Alterar os valores no arquivo <code>oozie-site.xml</code> do Oozie.	Restarts Oozie.

Classificações	Descrição	Ações de reconfiguração
phoenix-hbase-metrics	Alterar os valores no arquivo <code>hadoop-metrics2-hbase.properties</code> do Phoenix.	Not available.
phoenix-hbase-site	Alterar os valores no arquivo <code>hbase-site.xml</code> do Phoenix.	Not available.
phoenix-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Alterar os valores no arquivo <code>hadoop-metrics2-phoenix.properties</code> do Phoenix.	Not available.
pig-env	Alterar os valores no ambiente do Pig.	Not available.
pig-properties	Alterar os valores no arquivo <code>pig.properties</code> do Pig.	Restarts Oozie.
pig-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Pig.	Not available.
presto-log	Alterar os valores no arquivo <code>log.properties</code> do Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Alterar os valores no arquivo <code>config.properties</code> do Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Altere os valores no arquivo <code>password-authenticator.properties</code> do Presto.	Not available.
presto-env	Altere os valores no arquivo <code>presto-env.sh</code> do Presto.	Restarts Presto-Server (for PrestoDB)

Classificações	Descrição	Ações de reconfiguração
presto-node	Altere os valores no arquivo <code>node.properties</code> do Presto.	Not available.
presto-connector-blackhole	Alterar os valores no arquivo <code>blackhole.properties</code> do Presto.	Not available.
presto-connector-cassandra	Alterar os valores no arquivo <code>cassandra.properties</code> do Presto.	Not available.
presto-connector-hive	Alterar os valores no arquivo <code>hive.properties</code> do Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Alterar os valores no arquivo <code>jmx.properties</code> do Presto.	Not available.
presto-connector-kafka	Alterar os valores no arquivo <code>kafka.properties</code> do Presto.	Not available.
presto-connector-lakeformat ion	Alterar os valores no arquivo <code>lakeformation.properties</code> do Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Alterar os valores no arquivo <code>localfile.properties</code> do Presto.	Not available.
presto-connector-memory	Alterar os valores no arquivo <code>memory.properties</code> do Presto.	Not available.
presto-connector-mongodb	Alterar os valores no arquivo <code>mongodb.properties</code> do Presto.	Not available.
presto-connector-mysql	Alterar os valores no arquivo <code>mysql.properties</code> do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.	Not available.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.	Not available.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.	Not available.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.	Not available.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.	Not available.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.	Not available.
trino-log	Alterar os valores no arquivo log.properties do Trino.	Restarts Trino-Server (for Trino)
trino-config	Alterar os valores no arquivo config.properties do Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Alterar os valores no arquivo password-authenticator.properties do Trino.	Restarts Trino-Server (for Trino)
trino-env	Alterar os valores no arquivo trino-env.sh do Trino.	Restarts Trino-Server (for Trino)
trino-node	Alterar os valores no arquivo node.properties do Trino.	Not available.

Classificações	Descrição	Ações de reconfiguração
trino-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Trino.	Not available.
trino-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Trino.	Not available.
trino-connector-delta	Alterar os valores no arquivo delta.properties do Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Alterar os valores no arquivo hive.properties do Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Alterar os valores no arquivo exchange-manager.properties do Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Alterar os valores no arquivo iceberg.properties do Trino.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Alterar os valores no arquivo hudi.properties do Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Alterar os valores no arquivo jmx.properties do Trino.	Not available.
trino-connector-kafka	Alterar os valores no arquivo kafka.properties do Trino.	Not available.
trino-connector-localfile	Alterar os valores no arquivo localfile.properties do Trino.	Not available.
trino-connector-memory	Alterar os valores no arquivo memory.properties do Trino.	Not available.
trino-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Trino.	Not available.

Classificações	Descrição	Ações de reconfiguração
trino-connector-mysql	Alterar os valores no arquivo mysql.properties do Trino.	Not available.
trino-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Trino.	Not available.
trino-connector-raptor	Alterar os valores no arquivo raptor.properties do Trino.	Not available.
trino-connector-redis	Alterar os valores no arquivo redis.properties do Trino.	Not available.
trino-connector-redshift	Alterar os valores no arquivo redshift.properties do Trino.	Not available.
trino-connector-tpch	Alterar os valores no arquivo tpch.properties do Trino.	Not available.
trino-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Trino.	Not available.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.	Restarts Ranger KMS Server.
ranger-kms-logback	Altere os valores no arquivo kms-logback.xml do Ranger. KMS	Not available.

Classificações	Descrição	Ações de reconfiguração
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.	Not available.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Alterar os valores no ambiente do Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.	Not available.
spark-log4j2	Alterar os valores no arquivo log4j2.properties do Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Alterar os valores no ambiente do Sqoop.	Not available.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.	Not available.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.	Not available.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.	Restart Oozie and HiveServer2.

Classificações	Descrição	Ações de reconfiguração
yarn-env	Mude os valores no YARN ambiente.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.	Restarts Zeppelin.
zeppelin-site	Alterar as definições da configuração em zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.	Restarts Zookeeper server.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.	Restarts Zookeeper server.

Log de alterações 6.10.1

Log de alterações da versão 6.10.1 e das notas da versão

Data	Evento	Descrição
2023-08-30	Atualizar notas da versão	Adicionadas várias correções relacionadas ao ambiente de gerenciamento às notas da versão
2023-08-21	Publicação de documentos	Notas de lançamento do Amazon EMR 6.10.1 publicadas pela primeira vez
2023-08-16	Implantação concluída	Amazon EMR 6.10.1 totalmente implantado em todas as regiões suportadas
2023-08-04	Lançamento inicial	Amazon EMR 6.10.1 foi implantado pela primeira vez em regiões comerciais limitadas

Amazon EMR versão 6.10.0

Versões da aplicação 6.10.0

As seguintes aplicações são compatíveis com esta versão: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)

- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-6.10.0	emr-6.9.1	emr-6.9.0	emr-6.8.1
AWS SDK para 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	2.12.15	2.12.15	2.12.15	2.12.15
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
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

	emr-6.10.0	emr-6.9.1	emr-6.9.0	emr-6.8.1
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	0.10.2	0.10.2	0.10.2	0.9.2
Trino (PrestoSQL)	403	398	398	388
Zeppelin	0.10.1	0.10.1	0.10.1	0.10.1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

Notas da versão 6.10.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 6.10.0 da Amazon. As alterações são referentes à versão 6.9.0. Para obter informações sobre o cronograma da versão, consulte o [log de alterações](#).

Novos atributos

- O Amazon EMR 6.10.0 é compatível com 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 e PrestoDB 0.278.1.
- O Amazon EMR 6.10.0 inclui um conector Trino-Hudi nativo que fornece acesso de leitura aos dados nas tabelas Hudi. Você pode ativar o conector com `trino-cli --catalog hudi` e configurar o conector de acordo com suas necessidades com `trino-connector-hudi`. A integração nativa com a Amazon EMR significa que você não precisa mais usar `trino-connector-hive` para consultar tabelas Hudi. Para obter uma lista das configurações compatíveis com o novo conector, consulte a página do [conector do Hudi](#) na documentação do Trino.
- As EMR versões 6.10.0 e superiores da Amazon oferecem suporte à integração do Apache Zeppelin com o Apache Flink. Consulte [Usar trabalhos do Flink pelo Zeppelin no Amazon EMR](#) Para mais informações.

Problemas conhecidos

- O Hadoop 3.3.3 introduziu uma alteração em YARN ([YARN-9608](#)) que mantém os nós onde os contêineres eram executados em um estado de descomissionamento até que o aplicativo seja concluído. Essa alteração garante que dados locais, como dados embaralhados, não sejam perdidos e que você não precise executar o trabalho novamente. Essa abordagem também pode levar à subutilização de recursos em clusters com ou sem o ajuste de escala gerenciado habilitado.

Para contornar esse problema no Amazon EMR 6.10.0, você pode definir o valor de `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` to `false` in `yarn-site.xml` Nas EMR versões 6.11.0 e superiores da Amazon, bem como nas 6.8.1, 6.9.1 e 6.10.1, a configuração é definida como padrão para resolver esse problema. `false`

Alterações, melhorias e problemas resolvidos

- O Amazon EMR 6.10.0 remove a dependência `minimal-json.jar` da [integração do Amazon Redshift com o Apache Spark e adiciona automaticamente os jars relacionados ao Spark-Redshift](#) necessários ao caminho da classe executora para o Spark:, e `spark-redshift.jar` `spark-avro.jar` `RedshiftJDBC.jar`

- A versão 6.10.0 aprimora o daemon de gerenciamento de registros no cluster para monitorar pastas de log adicionais em seu cluster. Essa melhoria minimiza os cenários de utilização excessiva do disco.
- A versão 6.10.0 reinicia automaticamente o daemon de gerenciamento de logs no cluster quando ele é interrompido. Essa melhoria reduz o risco de os nós parecerem não íntegros devido à utilização excessiva do disco.
- O Amazon EMR 6.10.0 oferece suporte a endpoints regionais para EMRFS mapeamento de usuários.
- O tamanho padrão do volume raiz aumentou para 15 GB no Amazon EMR 6.10.0 e superior. O tamanho padrão do volume raiz das versões anteriores é de 10 GB.
- A versão 6.10.0 corrige um problema que fazia com que as tarefas do Spark parassem quando todos os executores restantes do Spark estavam em um host de desativação com o gerente de recursos. YARN
- Com o Amazon EMR 6.6.0 a 6.9.x, INSERT as consultas com partição dinâmica e uma cláusula ORDER BY ou SORT BY sempre terão dois redutores. Esse problema é causado pela OSS alteração [HIVE-20703](#), que coloca a otimização dinâmica da partição de classificação sob decisão baseada em custos. Se sua workload não exigir a classificação de partições dinâmicas, recomendamos que você defina a propriedade `hive.optimize.sort.dynamic.partition.threshold` como -1 para desabilitar o novo atributo e obter o número de redutores calculado corretamente. Esse problema foi corrigido no OSS Hive como parte do [HIVE-22269](#) e foi corrigido no Amazon 6.10.0. EMR
- Quando você inicia um cluster com a versão de patch mais recente do Amazon EMR 5.36 ou superior, 6.6 ou superior ou 7.0 ou superior, a Amazon EMR usa a versão mais recente do Amazon Linux 2023 ou Amazon Linux 2 para a Amazon padrão. EMR AMI Para obter mais informações, consulte [Usando o Amazon Linux padrão AMI para a Amazon EMR](#).

Note

Esta versão não recebe mais AMI atualizações automáticas, pois foi sucedida por mais 1 lançamento de patch. A versão de patch é indicada pelo número após o segundo ponto decimal (6.8.**1**). Para ver se você está usando a versão de patch mais recente, verifique as versões disponíveis no [Guia de lançamento](#) ou verifique o menu suspenso de EMRlançamentos da Amazon ao criar um cluster no console ou usar a ação [ListReleaseLabels](#)API ou [list-release-labels](#)CLI. Para receber atualizações sobre novos lançamentos, assine o RSS feed do [What's new?](#) página.

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 709.1	4.14.348	23 de julho de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia), Ásia

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Pacífico (Hyderabad), Oriente Médio (), Europa (EspanhaUAE), Europa (Zurique), Ásia-Pacífico (Melbourne), Israel (Tel Aviv), Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 808.0	4.14.320	24 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), Israel (Tel Aviv).

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 727.0	4.14.320	14 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 719.0	4.14.320	2 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)
2.0.2023 628.0	4.14.318	12 de julho de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Zurique), Europa (Milão), Europa (Espanha), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad)), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio () UAE

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 612.0	4.14.314	23 de junho de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Zurique), Europa (Milão), Europa (Espanha), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad)), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio () UAE

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 504.1	4.14.313	16 de maio de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Zurique), Europa (Milão), Europa (Espanha), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad)), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio () UAE

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 418.0	4.14.311	3 de maio de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Zurique), Europa (Milão), Europa (Espanha), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad)), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio () UAE

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 404.1	4.14.311	18 de abril de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul)), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio () UAE
2.0.2023 404.0	4.14.311	10 de abril de 2023	Leste dos EUA (Norte da Virgínia), Europa (Paris)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 320.0	4.14.309	30 de março de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul)), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio () UAE

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 207.0	4.14.304	22 de fevereiro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul)), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio () UAE

Versões de componente 6.10.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são

a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado myapp-component com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como 2.2-amzn-2.

Componente	Version (Versão)	Descrição
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark SDK
delta	2.2.0	O Delta Lake é um formato de tabela aberto para grandes conjuntos de dados analíticos
emr-ddb	4.16.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	3.3.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.7.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-notebook-env	1.7.0	Ambiente Conda para bloco de anotações do emr, que inclui o jupyter enterprise gateway
emr-s3-dist-cp	2.24.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	2.3.0	EMRConector S3Select

Componente	Version (Versão)	Descrição
emr-wal-cli	1.0.0	Cli usada para lista/exclusão de emrwal.
emrfs	2.55.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.16.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
flink-jobmanager-config	1.16.0	Gerenciando recursos em EMR nós para o Apache JobManager Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	3.3.3-amzn-2	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	3.3.3-amzn-2	HDFSserviço em nível de nó para armazenar blocos.

Componente	Version (Versão)	Descrição
hadoop-hdfs-library	3.3.3-amzn-2	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	3.3.3-amzn-2	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	3.3.3-amzn-2	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-https-server	3.3.3-amzn-2	HTTPendpoint para HDFS operações.
hadoop-kms-server	3.3.3-amzn-2	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	3.3.3-amzn-2	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	3.3.3-amzn-2	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	3.3.3-amzn-2	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	3.3.3-amzn-2	Serviço para recuperar informações atuais e históricas para YARN aplicativos.

Componente	Version (Versão)	Descrição
hbase-hmaster	2.4.15-amzn-0	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	2.4.15-amzn-0	Serviço para atender a uma ou mais HBase regiões.
hbase-client	2.4.15-amzn-0	HBase cliente de linha de comando.
hbase-rest-server	2.4.15-amzn-0	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	2.4.15-amzn-0	Serviço que fornece um endpoint Thrift para HBase.
hbase-operator-tools	2.4.15-amzn-0	Ferramenta de reparo para HBase clusters Apache.
hcatalog-client	3.1.3-amzn-3	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	3.1.3-amzn-3	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	3.1.3-amzn-3	HTTP endpoint fornecendo uma REST interface para HCatalog.

Componente	Version (Versão)	Descrição
hive-client	3.1.3-amzn-3	O cliente da linha de comando do Hive.
hive-hbase	3.1.3-amzn-3	Cliente do Hive-hbase.
hive-metastore-server	3.1.3-amzn-3	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	3.1.3-amzn-3	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.12.2-amzn-0	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.
hudi-presto	0.12.2-amzn-0	Biblioteca de pacotes para executar o Presto com o Hudi.
hudi-trino	0.12.2-amzn-0	Biblioteca de pacotes para execução do Presto com o Hudi.
hudi-spark	0.12.2-amzn-0	Biblioteca de pacotes para executar o Spark com o Hudi.
hue-server	4.10.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop

Componente	Version (Versão)	Descrição
iceberg	1.1.0-amzn-0	Apache Iceberg é um formato de tabela aberta para conjuntos de dados analíticos imensos
jupyterhub	1.5.0	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.7.1-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mxnet	1.9.1	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mariadb-server	5.5.68+	Servidor de banco de dados MariaDB.
nvidia-cuda	11.8.0	Drivers NVIDIA e toolkit CUDA
oozie-client	5.2.1	O cliente da linha de comando do Oozie.
oozie-server	5.2.1	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	4.5.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	5.1.2	As bibliotecas do phoenix para servidor e cliente

Componente	Version (Versão)	Descrição
phoenix-connectors	6.0.0- SNAPSHOT	Conectores do Apache Phoenix para Spark-3
phoenix-query-server	6.0.0	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.278.1-amzn-0	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.278.1-amzn-0	O serviço que executa partes de uma consulta.
presto-client	0.278.1-amzn-0	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
trino-coordinator	403-amzn-0	Serviço que aceita consultas e gerencia a execução de consultas entre trino-workers.
trino-worker	403-amzn-0	O serviço que executa partes de uma consulta.
trino-client	403-amzn-0	Cliente de linha de comando do Trino instalado em mestres de espera de um cluster de alta disponibilidade (HA) em que o servidor do Trino não é iniciado.

Componente	Version (Versão)	Descrição
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	4.0.2	O projeto R para computação estatística
ranger-kms-server	2.0.0	Sistema de gerenciamento de chaves Apache Ranger
spark-client	3.3.1-amzn-0	Os clientes da linha de comando do Spark.
spark-history-server	3.3.1-amzn-0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	3.3.1-amzn-0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	3.3.1-amzn-0	Bibliotecas do Apache Spark necessárias para escravos. YARN
spark-rapids	22.12.0-amzn-0	RAPIDSPlugin Nvidia Spark que acelera o Apache Spark com o. GPUs
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	2.11.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.

Componente	Version (Versão)	Descrição
tez-on-yarn	0.10.2-amzn-1	O YARN aplicativo e as bibliotecas tez.
tez-on-worker	0.10.2-amzn-1	O YARN aplicativo tez e as bibliotecas para nós de trabalho.
webserver	2.4.41+	HTTPServidor Apache.
zeppelin-server	0.10.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.5.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.5.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 6.10.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Ocorrem ações de reconfiguração quando você especifica uma configuração para grupos de instâncias em um cluster em execução. A Amazon EMR só inicia ações de reconfiguração para as classificações que você modifica. Para ter mais informações, consulte [Reconfigurar um grupo de instâncias de um cluster em execução](#).

Classificações do emr-6.10.0

Classificações	Descrição	Ações de reconfiguração
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.	Restarts the ResourceManager service.
container-executor	Altere os valores no arquivo container-executor.cfg YARN do Hadoop.	Not available.
container-log4j	Altere os valores no arquivo container-log4j.properties YARN do Hadoop.	Not available.
core-site	Alterar os valores no arquivo core-site.xml do 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	Alterar as configurações relacionadas ao docker.	Not available.
emrfs-site	Altere EMRFS as configurações.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

Classificações	Descrição	Ações de reconfiguração
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Alterar as configurações do flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Alterar as configurações de log4j.properties no Flink.	Restarts Flink history server.
flink-log4j-session	Alterar as configurações de log4j-yarn-session.properties no Flink para sessão do Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.	Restarts Flink history server.

Classificações	Descrição	Ações de reconfiguração
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do 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	Alterar os valores no arquivo log4j.properties do 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	Alterar a configuração do servidor SSL no Hadoop	Not available.
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop	Not available.

Classificações	Descrição	Ações de reconfiguração
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Mude os valores no HBase ambiente.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.	Not available.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.	This classification should not be reconfigured.
hdfs-env	Mude os valores no HDFS ambiente.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classificações	Descrição	Ações de reconfiguração
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Mude os valores no HCatalog ambiente.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Configurações EMR selecionadas pela Amazon para o Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.	Not available.

Classificações	Descrição	Ações de reconfiguração
hive-parquet-logging	Alterar os valores no arquivo <code>parquet-logging.properties</code> do Hive.	Not available.
hive-env	Alterar os valores no ambiente do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Altere os valores no arquivo <code>hive-exec-log 4j2.properties</code> do Hive.	Not available.
hive-llap-daemon-log4j2	Altere os valores no arquivo <code>llap-daemon-log 4j2.properties</code> do Hive.	Not available.
hive-log4j2	Alterar os valores no arquivo <code>hive-log4j2.properties</code> do Hive.	Not available.
hive-site	Alterar os valores no arquivo <code>hive-site.xml</code> do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Alterar os valores no arquivo <code>hiveserver2-site.xml</code> do Hive Server2.	Not available.
hue-ini	Alterar os valores no arquivo <code>ini</code> do Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.

Classificações	Descrição	Ações de reconfiguração
httpfs-env	Mude os valores no HTTPFS ambiente.	Restarts Hadoop Httpfs service.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.	Not available.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.	Not available.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Altere os valores no ambiente do Hudi.	Not available.
hudi-defaults	Alterar os valores no arquivo hudi-defaults.conf do Hudi.	Not available.
iceberg-defaults	Alterar os valores no arquivo iceberg-defaults.conf do Iceberg.	Not available.
delta-defaults	Alterar os valores no arquivo delta-defaults.conf do Delta.	Not available.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.	Not available.

Classificações	Descrição	Ações de reconfiguração
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo <code>jupyterhub_config.py</code> do.	Not available.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.	Not available.
jupyter-sparkmagic-conf	Altere os valores no arquivo <code>config.json</code> do Sparkmagic.	Not available.
livy-conf	Alterar os valores no arquivo <code>livy.conf</code> do Livy.	Restarts Livy Server.
livy-env	Alterar os valores no ambiente do Livy.	Restarts Livy Server.
livy-log4j2	Alterar as configurações de <code>log4j2.properties</code> do Livy.	Restarts Livy Server.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Altere os valores no arquivo <code>mapred-site.xml</code> do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Alterar os valores no ambiente do Oozie.	Restarts Oozie.
oozie-log4j	Alterar os valores no arquivo <code>oozie-log4j.properties</code> do Oozie.	Restarts Oozie.
oozie-site	Alterar os valores no arquivo <code>oozie-site.xml</code> do Oozie.	Restarts Oozie.

Classificações	Descrição	Ações de reconfiguração
phoenix-hbase-metrics	Alterar os valores no arquivo <code>hadoop-metrics2-hbase.properties</code> do Phoenix.	Not available.
phoenix-hbase-site	Alterar os valores no arquivo <code>hbase-site.xml</code> do Phoenix.	Not available.
phoenix-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Alterar os valores no arquivo <code>hadoop-metrics2-phoenix.properties</code> do Phoenix.	Not available.
pig-env	Alterar os valores no ambiente do Pig.	Not available.
pig-properties	Alterar os valores no arquivo <code>pig.properties</code> do Pig.	Restarts Oozie.
pig-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Pig.	Not available.
presto-log	Alterar os valores no arquivo <code>log.properties</code> do Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Alterar os valores no arquivo <code>config.properties</code> do Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Altere os valores no arquivo <code>password-authenticator.properties</code> do Presto.	Not available.
presto-env	Altere os valores no arquivo <code>presto-env.sh</code> do Presto.	Restarts Presto-Server (for PrestoDB)

Classificações	Descrição	Ações de reconfiguração
presto-node	Altere os valores no arquivo <code>node.properties</code> do Presto.	Not available.
presto-connector-blackhole	Alterar os valores no arquivo <code>blackhole.properties</code> do Presto.	Not available.
presto-connector-cassandra	Alterar os valores no arquivo <code>cassandra.properties</code> do Presto.	Not available.
presto-connector-hive	Alterar os valores no arquivo <code>hive.properties</code> do Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Alterar os valores no arquivo <code>jmx.properties</code> do Presto.	Not available.
presto-connector-kafka	Alterar os valores no arquivo <code>kafka.properties</code> do Presto.	Not available.
presto-connector-lakeformat ion	Alterar os valores no arquivo <code>lakeformation.properties</code> do Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-localfile	Alterar os valores no arquivo <code>localfile.properties</code> do Presto.	Not available.
presto-connector-memory	Alterar os valores no arquivo <code>memory.properties</code> do Presto.	Not available.
presto-connector-mongodb	Alterar os valores no arquivo <code>mongodb.properties</code> do Presto.	Not available.
presto-connector-mysql	Alterar os valores no arquivo <code>mysql.properties</code> do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.	Not available.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.	Not available.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.	Not available.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.	Not available.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.	Not available.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.	Not available.
trino-log	Alterar os valores no arquivo log.properties do Trino.	Restarts Trino-Server (for Trino)
trino-config	Alterar os valores no arquivo config.properties do Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Alterar os valores no arquivo password-authenticator.properties do Trino.	Restarts Trino-Server (for Trino)
trino-env	Alterar os valores no arquivo trino-env.sh do Trino.	Restarts Trino-Server (for Trino)
trino-node	Alterar os valores no arquivo node.properties do Trino.	Not available.

Classificações	Descrição	Ações de reconfiguração
trino-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Trino.	Not available.
trino-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Trino.	Not available.
trino-connector-delta	Alterar os valores no arquivo delta.properties do Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Alterar os valores no arquivo hive.properties do Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Alterar os valores no arquivo exchange-manager.properties do Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Alterar os valores no arquivo iceberg.properties do Trino.	Restarts Trino-Server (for Trino)
trino-connector-hudi	Alterar os valores no arquivo hudi.properties do Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Alterar os valores no arquivo jmx.properties do Trino.	Not available.
trino-connector-kafka	Alterar os valores no arquivo kafka.properties do Trino.	Not available.
trino-connector-localfile	Alterar os valores no arquivo localfile.properties do Trino.	Not available.
trino-connector-memory	Alterar os valores no arquivo memory.properties do Trino.	Not available.
trino-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Trino.	Not available.

Classificações	Descrição	Ações de reconfiguração
trino-connector-mysql	Alterar os valores no arquivo mysql.properties do Trino.	Not available.
trino-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Trino.	Not available.
trino-connector-raptor	Alterar os valores no arquivo raptor.properties do Trino.	Not available.
trino-connector-redis	Alterar os valores no arquivo redis.properties do Trino.	Not available.
trino-connector-redshift	Alterar os valores no arquivo redshift.properties do Trino.	Not available.
trino-connector-tpch	Alterar os valores no arquivo tpch.properties do Trino.	Not available.
trino-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Trino.	Not available.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.	Restarts Ranger KMS Server.
ranger-kms-logback	Altere os valores no arquivo kms-logback.xml do Ranger. KMS	Not available.

Classificações	Descrição	Ações de reconfiguração
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.	Not available.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Alterar os valores no ambiente do Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.	Not available.
spark-log4j2	Alterar os valores no arquivo log4j2.properties do Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Alterar os valores no ambiente do Sqoop.	Not available.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.	Not available.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.	Not available.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.	Restart Oozie and HiveServer2.

Classificações	Descrição	Ações de reconfiguração
yarn-env	Mude os valores no YARN ambiente.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.	Restarts Zeppelin.
zeppelin-site	Alterar as definições da configuração em zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.	Restarts Zookeeper server.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.	Restarts Zookeeper server.

Log de alterações 6.10.0

Log de alterações da versão 6.10.0 e das notas da versão

Data	Evento	Descrição
2023-08-21	Atualizar	Adicionado um problema conhecido introduzido com o Hadoop 3.3.3.
2023-07-26	Atualizar	Novos rótulos de versão 2.0.20230612.0 e 2.0.20230628.0 do sistema operacional.
2023-03-02	Implantação concluída	Amazon EMR 6.10 totalmente implantado em todas as regiões suportadas
2023-03-02	Publicação de documentos	Notas de lançamento do Amazon EMR 6.10 publicadas pela primeira vez
2023-02-27	Lançamento inicial	Amazon EMR 6.10 implantado em regiões comerciais limitadas

Amazon EMR versão 6.9.1

Versões 6.9.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-6.9.1	emr-6.9.0	emr-6.8.1	emr-6.8.0
AWS SDK para 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	2.12.15	2.12.15	2.12.15	2.12.15
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
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

	emr-6.9.1	emr-6.9.0	emr-6.8.1	emr-6.8.0
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	0.10.2	0.10.2	0.9.2	0.9.2
Trino (PrestoSQL)	398	398	388	388
Zeppelin	0.10.1	0.10.1	0.10.1	0.10.1
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.10

Notas da versão 6.9.1

As notas de lançamento a seguir incluem informações sobre a EMR versão 6.9.1 da Amazon. As alterações são referentes à versão 6.9.0. Para obter informações sobre o cronograma da versão, consulte o [Log de alterações 6.9.1](#).

Alterações, melhorias e problemas resolvidos

- O Hadoop 3.3.3 introduziu uma alteração em YARN ([YARN-9608](#)) que mantém os nós onde os contêineres eram executados em um estado de descomissionamento até que o aplicativo seja concluído. Essa alteração garante que dados locais, como dados embaralhados, não sejam perdidos e que você não precise executar o trabalho novamente. Essa abordagem também pode levar à subutilização de recursos em clusters com ou sem o ajuste de escala gerenciado habilitado.

Com as EMR versões 6.11.0 e superiores da Amazon, bem como 6.8.1, 6.9.1 e 6.10.1, o valor de `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` é definido como `in` para resolver esse problema. `false yarn-site.xml`

Embora a correção resolva os problemas introduzidos pelo YARN -9608, ela pode fazer com que as tarefas do Hive falhem devido à perda aleatória de dados em clusters com escalabilidade gerenciada habilitada. Reduzimos esse risco nesta versão também ao configurar `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-shuffle-data` para workloads do Hive. Essa configuração só está disponível nas EMR versões 6.11.0 e superiores da Amazon.

- O coletor de métricas não enviará métricas para o ambiente de gerenciamento após o failover do nó primário em clusters com a configuração de grupos de instâncias.
- Essa versão elimina novas tentativas de HTTP solicitações malsucedidas para endpoints do coletor de métricas.
- Esta versão inclui uma alteração que permite que clusters de alta disponibilidade se recuperem de um estado de falha após a reinicialização.
- Esta versão corrige um problema em que grandes exceções de estouro criadas por usuários causavam exceções de estouro.
- Esta versão corrige problemas de tempo limite com o processo de EMR reconfiguração da Amazon.
- Esta versão inclui correções de segurança.

- Esta versão corrige um problema em que clusters que estão executando cargas de trabalho no Spark com a Amazon EMR podem receber silenciosamente resultados incorretos com `contains`, `startsWith`, e `endsWith` `like`. Esse problema ocorre quando você usa as expressões em campos particionados que têm metadados no Amazon EMR Hive3 Metastore Server (`.`). HMS
- Com o Amazon EMR 6.6.0 a 6.9.x, `INSERT` as consultas com partição dinâmica e uma cláusula `ORDER BY` ou `SORT BY` sempre terão dois redutores. Esse problema é causado pela OSS alteração [HIVE-20703](#), que coloca a otimização dinâmica da partição de classificação sob decisão baseada em custos. Se sua workload não exigir a classificação de partições dinâmicas, recomendamos que você defina a propriedade `hive.optimize.sort.dynamic.partition.threshold` como `-1` para desabilitar o novo atributo e obter o número de redutores calculado corretamente. Esse problema foi corrigido no OSS Hive como parte do [HIVE-22269](#) e foi corrigido no Amazon 6.10.0. EMR
- O Hive pode sofrer perda de dados quando você o usa HDFS como um diretório temporário e habilita a mesclagem de arquivos pequenos, e a tabela contém caminhos de partição estáticos.
- Esta versão corrige um problema de desempenho com o Hive se a mesclagem de arquivos pequenos (desativada por padrão) estiver ativada no final do ETL trabalho.
- Esta versão corrige um problema de limitação no lado do Glue quando não há funções definidas pelo usuário (`.`). UDF
- Esta versão corrige um problema que exclui registros de contêineres pelo serviço de agregação de registros de nós antes que o carregador de registros possa enviá-los para o S3 em caso de descomissionamento. YARN
- Esta versão corrige o manuseio de arquivos compactados/arquivados com rastreamento persistente de arquivos de armazenamento para. HBase
- Esta versão corrige um problema que afetou a performance do Spark quando você definiu um valor de `true` padrão para a configuração `spark.yarn.heterogeneousExecutors.enabled` no `spark-defaults.conf`.
- Esta versão corrige um problema com a falha do Reduce Task em ler dados embaralhados. O problema causou falhas na consulta do Hive com um erro de memória corrompida.
- Esta versão corrige um problema que fazia com que o provisionador de nós falhasse se o serviço HDFS NameNode (NN) ficasse preso no modo de segurança durante a substituição do nó.
- Esta versão adiciona um novo mecanismo de repetição ao fluxo de trabalho de escalabilidade de EMR clusters para clusters que executam o Presto ou o Trino. Essa melhoria reduz o risco de que o redimensionamento do cluster fique paralisado indefinidamente devido a uma única falha

na operação de redimensionamento. Ela também aprimora a utilização dos clusters, porque seu cluster aumenta e reduz a escala verticalmente com mais rapidez.

- Esta versão melhora a lógica de redução do cluster para que seu cluster não tente reduzir os nós principais abaixo da configuração do fator de HDFS replicação do cluster. Isso se alinha aos seus requisitos de redundância de dados e reduz a probabilidade de uma operação de escalabilidade paralisar.
- O daemon de gerenciamento de logs foi atualizado para identificar todos os logs que estão em uso ativo com identificadores de arquivos abertos no armazenamento da instância local e nos processos associados. Essa atualização garante que a Amazon exclua EMR adequadamente os arquivos e recupere o espaço de armazenamento depois que os registros forem arquivados no Amazon S3.
- Esta versão inclui um aprimoramento do daemon de gerenciamento de logs que exclui diretórios de etapas vazios e não utilizados no sistema de arquivos de cluster local. Um número excessivamente grande de diretórios vazios pode degradar o desempenho dos EMR daemons da Amazon e resultar na sobreutilização do disco.
- Esta versão corrige um problema que pode ocorrer quando você cria um nó de borda ao replicar um dos nós primários de um cluster com vários nós primários. O nó de borda replicado pode causar atrasos nas operações de redução da escala verticalmente ou resultar em alta utilização de memória nos nós primários. Para obter mais informações sobre como criar um nó de borda para se comunicar com seu EMR cluster, consulte [Edge Node Creator](#) no `aws-samples` repositório em GitHub.
- Esta versão melhora o processo de automação que a Amazon EMR usa para remontar EBS volumes da Amazon em uma instância após uma reinicialização.
- Esta versão corrige um problema que resultou em lacunas intermitentes nas métricas do Hadoop que a Amazon publica EMR na Amazon. CloudWatch
- Esta versão corrige um problema com EMR clusters em que uma atualização no arquivo de YARN configuração que contém a lista de exclusão de nós do cluster é interrompida devido à utilização excessiva do disco. A atualização incompleta impede futuras operações de redução da escala verticalmente do cluster. Esta versão garante que o cluster permaneça íntegro e que as operações de escalabilidade funcionem conforme esperado.
- Esta versão aprimora o daemon de gerenciamento de logs no cluster para monitorar pastas de log adicionais em seu cluster. EMR Essa melhoria minimiza os cenários de utilização excessiva do disco.

- Esta versão reinicia automaticamente o daemon de gerenciamento de logs no cluster quando ele é interrompido. Essa melhoria reduz o risco de os nós parecerem não íntegros devido à utilização excessiva do disco.
- Quando você inicia um cluster com a versão de patch mais recente do Amazon EMR 5.36 ou superior, 6.6 ou superior ou 7.0 ou superior, a Amazon EMR usa a versão mais recente do Amazon Linux 2023 ou Amazon Linux 2 para a Amazon padrão. EMR AMI Para obter mais informações, consulte [Usando o Amazon Linux padrão AMI para a Amazon EMR](#).

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 709.1	4.14.348	23 de julho de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia), Ásia Pacífico (Hyderabad), Oriente Médio (), Europa (EspanhaUAE), Europa (Zurique), Ásia-Pacífico (Melbourne), Israel (Tel Aviv), Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 223.0	4.14.336	8 de março de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Canadá (Central), Israel (Tel Aviv), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			China (Ningxia), Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 131.0	4.14.336	14 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Canadá (Central), Israel (Tel Aviv), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			China (Ningxia), Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 124.0	4.14.336	7 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Canadá (Central), Israel (Tel Aviv), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			China (Ningxia), Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 109.0	4.14.334	24 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Canadá (Central), Israel (Tel Aviv), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim),

OsRelease Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			China (Ningxia), Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 218.0	4.14.330	2 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Canadá (Central), Israel (Tel Aviv), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 206.0	4.14.330	22 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Canadá (Central), Israel (Tel Aviv), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 116.0	4.14.328	11 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Canadá (Central), Israel (Tel Aviv), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 101.0	4.14.327	16 de novembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Canadá (Central), Israel (Tel Aviv), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 020.1	4.14.326	7 de novembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Canadá (Central), Israel (Tel Aviv), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 012.1	4.14.326	26 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Canadá (Central), Israel (Tel Aviv), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 926.0	4.14.322	19 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Canadá (Central), Israel (Tel Aviv), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 906.0	4.14.322	4 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Canadá (Central), Israel (Tel Aviv).

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 822.0	4.14.322	30 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Canadá (Central), Israel (Tel Aviv).

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 808.0	4.14.320	24 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Canadá (Central), Israel (Tel Aviv).

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 727.0	4.14.320	14 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central)

Versões de componente 6.9.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são

a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado myapp-component com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como 2.2-amzn-2.

Componente	Version (Versão)	Descrição
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark SDK
delta	2.1.0	O Delta Lake é um formato de tabela aberto para grandes conjuntos de dados analíticos
emr-ddb	4.16.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	3.3.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.6.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-notebook-env	1.7.0	Ambiente Conda para bloco de anotações do emr, que inclui o jupyter enterprise gateway
emr-s3-dist-cp	2.23.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	2.2.0	EMRConector S3Select

Componente	Version (Versão)	Descrição
emrfs	2.54.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.15.2	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
flink-jobmanager-config	1.15.2	Gerenciando recursos em EMR nós para o Apache JobManager Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	3.3.3-amzn-1.1	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	3.3.3-amzn-1.1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	3.3.3-amzn-1.1	HDFScliente e biblioteca de linha de comando

Componente	Version (Versão)	Descrição
hadoop-hdfs-namenode	3.3.3-amzn-1.1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	3.3.3-amzn-1.1	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httpfs-server	3.3.3-amzn-1.1	HTTPendpoint para HDFS operações.
hadoop-kms-server	3.3.3-amzn-1.1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	3.3.3-amzn-1.1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	3.3.3-amzn-1.1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	3.3.3-amzn-1.1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	3.3.3-amzn-1.1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	2.4.13-amzn-0.1	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.

Componente	Version (Versão)	Descrição
hbase-region-server	2.4.13-amzn-0.1	Serviço para atender a uma ou mais HBase regiões.
hbase-client	2.4.13-amzn-0.1	HBase cliente de linha de comando.
hbase-rest-server	2.4.13-amzn-0.1	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	2.4.13-amzn-0.1	Serviço que fornece um endpoint Thrift para HBase
hbase-operator-tools	2.4.13-amzn-0.1	Ferramenta de reparo para HBase clusters Apache.
hcatalog-client	3.1.3-amzn-2.1	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	3.1.3-amzn-2.1	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	3.1.3-amzn-2.1	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	3.1.3-amzn-2.1	O cliente da linha de comando do Hive.
hive-hbase	3.1.3-amzn-2.1	Cliente do Hive-hbase.

Componente	Version (Versão)	Descrição
hive-metastore-server	3.1.3-amzn-2.1	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	3.1.3-amzn-2.1	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.12.1-amzn-0	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.
hudi-presto	0.12.1-amzn-0	Biblioteca de pacotes para executar o Presto com o Hudi.
hudi-trino	0.12.1-amzn-0	Biblioteca de pacotes para execução do Presto com o Hudi.
hudi-spark	0.12.1-amzn-0	Biblioteca de pacotes para executar o Spark com o Hudi.
hue-server	4.10.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
iceberg	0.14.1-amzn-0	Apache Iceberg é um formato de tabela aberta para conjuntos de dados analíticos imensos
jupyterhub	1.4.1	Servidor de vários usuários para blocos de anotações Jupyter

Componente	Version (Versão)	Descrição
livy-server	0.7.1-incubating	REST interface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mxnet	1.9.1	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mariadb-server	5.5.68+	Servidor de banco de dados MariaDB.
nvidia-cuda	11.7.0	Drivers NVIDIA e toolkit CUDA
oozie-client	5.2.1	O cliente da linha de comando do Oozie.
oozie-server	5.2.1	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	4.5.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	5.1.2	As bibliotecas do phoenix para servidor e cliente
phoenix-connectors	6.0.0- SNAPSHOT	Conectores do Apache Phoenix para Spark-3
phoenix-query-server	6.0.0	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API

Componente	Version (Versão)	Descrição
presto-coordinator	0.276-amzn-0	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.276-amzn-0	O serviço que executa partes de uma consulta.
presto-client	0.276-amzn-0	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
trino-coordinator	398-amzn-0	Serviço que aceita consultas e gerencia a execução de consultas entre trino-workers.
trino-worker	398-amzn-0	O serviço que executa partes de uma consulta.
trino-client	398-amzn-0	Cliente de linha de comando do Trino instalado em mestres de espera de um cluster de alta disponibilidade (HA) em que o servidor do Trino não é iniciado.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	4.0.2	O projeto R para computação estatística
ranger-kms-server	2.0.0	Sistema de gerenciamento de chaves Apache Ranger

Componente	Version (Versão)	Descrição
spark-client	3.3.0-amzn-1.1	Os clientes da linha de comando do Spark.
spark-history-server	3.3.0-amzn-1.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	3.3.0-amzn-1.1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	3.3.0-amzn-1.1	Bibliotecas do Apache Spark necessárias para escravos. YARN
spark-rapids	22.08.0-amzn-0	RAPIDSPlugin Nvidia Spark que acelera o Apache Spark com o. GPUs
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	2.10.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.10.2-amzn-0.1	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.41+	HTTP Servidor Apache.
zeppelin-server	0.10.1	O notebook baseado na web que permite a análise de dados interativa.

Componente	Version (Versão)	Descrição
zookeeper-server	3.5.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.5.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 6.9.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Ocorrem ações de reconfiguração quando você especifica uma configuração para grupos de instâncias em um cluster em execução. A Amazon EMR só inicia ações de reconfiguração para as classificações que você modifica. Para ter mais informações, consulte [Reconfigurar um grupo de instâncias de um cluster em execução](#).

Classificações do emr-6.9.1

Classificações	Descrição	Ações de reconfiguração
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.	Restarts the Resource Manager service.
container-executor	Altere os valores no arquivo <code>container-executor.cfg</code> YARN do Hadoop.	Not available.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.	Not available.

Classificações	Descrição	Ações de reconfiguração
core-site	Alterar os valores no arquivo core-site.xml do 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	Alterar as configurações relacionadas ao docker.	Not available.
emrfs-site	Altere EMRFS as configurações.	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.

Classificações	Descrição	Ações de reconfiguração
flink-conf	Alterar as configurações do flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Alterar as configurações de log4j.properties no Flink.	Restarts Flink history server.
flink-log4j-session	Alterar as configurações de log4j-yarn-session.properties no Flink para sessão do Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.	Restarts Flink history server.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do 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.

Classificações	Descrição	Ações de reconfiguração
hadoop-log4j	Alterar os valores no arquivo log4j.properties do 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	Alterar a configuração do servidor SSL no Hadoop	Not available.
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop	Not available.
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Mude os valores no HBase ambiente.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.

Classificações	Descrição	Ações de reconfiguração
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.	Not available.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.	This classification should not be reconfigured.
hdfs-env	Mude os valores no HDFS ambiente.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Mude os valores no HCatalog ambiente.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.	Restarts Hive WebHCat server.

Classificações	Descrição	Ações de reconfiguração
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Configurações EMR selecionadas pela Amazon para o Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.	Not available.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.	Not available.
hive-env	Alterar os valores no ambiente do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log 4j2.properties do Hive.	Not available.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log 4j2.properties do Hive.	Not available.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.	Not available.

Classificações	Descrição	Ações de reconfiguração
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.	Not available.
hue-ini	Alterar os valores no arquivo ini do Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
https-env	Mude os valores no HTTPFS ambiente.	Restarts Hadoop Httpfs service.
https-site	Alterar os valores no arquivo https-site.xml do Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.	Not available.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.	Not available.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Altere os valores no ambiente do Hudi.	Not available.

Classificações	Descrição	Ações de reconfiguração
hudi-defaults	Alterar os valores no arquivo hudi-defaults.conf do Hudi.	Not available.
iceberg-defaults	Alterar os valores no arquivo iceberg-defaults.conf do Iceberg.	Not available.
delta-defaults	Alterar os valores no arquivo delta-defaults.conf do Delta.	Not available.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.	Not available.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.	Not available.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.	Not available.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.	Not available.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.	Restarts Livy Server.
livy-env	Alterar os valores no ambiente do Livy.	Restarts Livy Server.
livy-log4j2	Alterar as configurações de log4j2.properties do Livy.	Restarts Livy Server.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.

Classificações	Descrição	Ações de reconfiguração
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Alterar os valores no ambiente do Oozie.	Restarts Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.	Restarts Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.	Not available.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.	Not available.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.	Not available.
pig-env	Alterar os valores no ambiente do Pig.	Not available.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.	Restarts Oozie.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-log	Alterar os valores no arquivo log.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Alterar os valores no arquivo config.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.	Not available.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Altere os valores no arquivo node.properties do Presto.	Not available.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.	Not available.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.	Not available.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.	Not available.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.	Not available.
presto-connector-lakeformation	Alterar os valores no arquivo lakeformation.properties do Presto.	Restarts Presto-Server (for PrestoDB)

Classificações	Descrição	Ações de reconfiguração
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.	Not available.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.	Not available.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.	Not available.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.	Not available.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.	Not available.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.	Not available.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.	Not available.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.	Not available.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.	Not available.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.	Not available.
trino-log	Alterar os valores no arquivo log.properties do Trino.	Restarts Trino-Server (for Trino)

Classificações	Descrição	Ações de reconfiguração
trino-config	Alterar os valores no arquivo <code>config.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Alterar os valores no arquivo <code>password-authenticator.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-env	Alterar os valores no arquivo <code>trino-env.sh</code> do Trino.	Restarts Trino-Server (for Trino)
trino-node	Alterar os valores no arquivo <code>node.properties</code> do Trino.	Not available.
trino-connector-blackhole	Alterar os valores no arquivo <code>blackhole.properties</code> do Trino.	Not available.
trino-connector-cassandra	Alterar os valores no arquivo <code>cassandra.properties</code> do Trino.	Not available.
trino-connector-delta	Alterar os valores no arquivo <code>delta.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Alterar os valores no arquivo <code>hive.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Alterar os valores no arquivo <code>exchange-manager.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Alterar os valores no arquivo <code>iceberg.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Alterar os valores no arquivo <code>jmx.properties</code> do Trino.	Not available.
trino-connector-kafka	Alterar os valores no arquivo <code>kafka.properties</code> do Trino.	Not available.

Classificações	Descrição	Ações de reconfiguração
trino-connector-localfile	Alterar os valores no arquivo localfile.properties do Trino.	Not available.
trino-connector-memory	Alterar os valores no arquivo memory.properties do Trino.	Not available.
trino-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Trino.	Not available.
trino-connector-mysql	Alterar os valores no arquivo mysql.properties do Trino.	Not available.
trino-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Trino.	Not available.
trino-connector-raptor	Alterar os valores no arquivo raptor.properties do Trino.	Not available.
trino-connector-redis	Alterar os valores no arquivo redis.properties do Trino.	Not available.
trino-connector-redshift	Alterar os valores no arquivo redshift.properties do Trino.	Not available.
trino-connector-tpch	Alterar os valores no arquivo tpch.properties do Trino.	Not available.
trino-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Trino.	Not available.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.	Restarts Ranger KMS Server.

Classificações	Descrição	Ações de reconfiguração
ranger-kms-env	Altere os valores no KMS ambiente Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Altere os valores no arquivo kms-log4j.properties do Ranger. KMS	Not available.
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.	Not available.
spark	Configurações EMR seleciona das pela Amazon para o Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Alterar os valores no ambiente do Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.	Not available.
spark-log4j2	Alterar os valores no arquivo log4j2.properties do Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Alterar os valores no ambiente do Sqoop.	Not available.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.	Not available.

Classificações	Descrição	Ações de reconfiguração
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.	Not available.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.	Restart Oozie and HiveServer2.
yarn-env	Mude os valores no YARN ambiente.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.	Restarts Zeppelin.
zeppelin-site	Alterar as definições da configuração em zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.	Restarts Zookeeper server.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.	Restarts Zookeeper server.

Log de alterações 6.9.1

Log de alterações da versão 6.9.1 e das notas da versão

Data	Evento	Descrição
2023-08-30	Atualizar notas da versão	Adicionadas várias correções relacionadas ao ambiente de gerenciamento às notas da versão
2023-08-21	Publicação de documentos	Notas de lançamento da Amazon EMR 6.9.1 publicadas pela primeira vez
2023-08-16	Implantação concluída	Amazon EMR 6.9.1 totalmente implantado em todas as regiões suportadas
2023-08-04	Lançamento inicial	Amazon EMR 6.9.1 foi implantado pela primeira vez em regiões comerciais limitadas

Amazon EMR versão 6.9.0

Versões da aplicação 6.9.0

As seguintes aplicações são compatíveis com esta versão: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)

- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-6.9.0	emr-6.8.1	emr-6.8.0	emr-6.7.0
AWS SDK para 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	2.12.15	2.12.15	2.12.15	2.12.15
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
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

	emr-6.9.0	emr-6.8.1	emr-6.8.0	emr-6.7.0
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	0.10.2	0.9.2	0.9.2	0.9.2
Trino (PrestoSQL)	398	388	388	378
Zeppelin	0.10.1	0.10.1	0.10.1	0.10.0
ZooKeeper	3.5.10	3.5.10	3.5.10	3.5.7

Notas da versão 6.9.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 6.9.0 da Amazon. As alterações são relativas à EMR versão 6.8.0 da Amazon. Para obter informações sobre o cronograma da versão, consulte o [log de alterações](#).

Novos atributos

- A EMR versão 6.9.0 da Amazon é compatível com Apache Spark RAPIDS 22.08.0, Apache Hudi 0.12.1, Apache Iceberg 0.14.1, Trino 398 e Tez 0.10.2.
- A EMR versão 6.9.0 da Amazon inclui um novo aplicativo de código aberto, 2.1.0. [Delta Lake](#)
- A integração do Amazon Redshift para o Apache Spark está incluída nas EMR versões 6.9.0 e posteriores da Amazon. Anteriormente uma ferramenta de código aberto, a integração nativa é um conector do Spark que você pode usar para criar aplicações do Apache Spark que realizam a leitura e a gravação de dados no Amazon Redshift e no Amazon Redshift sem servidor. Para ter mais informações, consulte [Usando a integração do Amazon Redshift para o Apache Spark com a Amazon EMR](#).
- A EMR versão 6.9.0 da Amazon adiciona suporte ao arquivamento de registros no Amazon S3 durante a redução da escala do cluster. Anteriormente, só era possível arquivar arquivos de log no Amazon S3 durante o encerramento do cluster. A nova capacidade garante que os arquivos de log gerados no cluster persistam no Amazon S3 mesmo após o encerramento do nó. Para obter mais informações, consulte [Configurar registro em log e depuração do cluster](#).
- Para dar suporte a consultas de longa execução, o Trino agora inclui um mecanismo de execução tolerante a falhas. A execução tolerante a falhas atenua as falhas nas consultas ao tentar novamente as consultas com falha ou as tarefas dos seus componentes. Para ter mais informações, consulte [Execução tolerante a falhas no Trino](#).
- Você pode usar o Apache Flink na Amazon EMR para unificar BATCH e STREAM processar tabelas do Apache Hive ou metadados de qualquer fonte de tabela do Flink, como Iceberg, Kinesis ou Kafka. Você pode especificar o AWS Glue Data Catalog como metastore para o Flink usando o AWS Management Console,, AWS CLI ou Amazon. EMR API Para ter mais informações, consulte [Configurando o Flink na Amazon EMR](#).
- Agora você pode especificar AWS Identity and Access Management (IAM) funções de tempo de execução e controle de acesso AWS Lake Formation baseado para consultas do Apache Spark, Apache Hive e Presto na Amazon EMR em EC2 clusters com o Amazon Studio. SageMaker Para obter mais informações, consulte [Configurar funções de tempo de execução para EMR as etapas da Amazon](#).

Problemas conhecidos

- Para a EMR versão 6.9.0 da Amazon, o Trino não funciona em clusters habilitados para o Apache Ranger. Se você precisar usar o Trino com o Ranger, entre em contato com o [AWS Support](#).

- Se você usar a integração do Amazon Redshift para Apache Spark e tiver um time, timetz, timestamp ou timestamptz com precisão de microssegundos no formato Parquet, o conector arredondará os valores de tempo para o valor de milissegundo mais próximo. Como solução alternativa, use o parâmetro `unload_s3_format` do formato de descarregamento de texto.
- Quando você usa o Spark com a formatação de localização de partições do Hive para ler dados no Amazon S3 e executa o Spark nas EMR versões 5.30.0 a 5.36.0 e 6.2.0 a 6.9.0 da Amazon, você pode encontrar um problema que impede seu cluster de ler os dados corretamente. Isso poderá acontecer se suas partições tiverem todas as características a seguir:
 - Duas ou mais partições são verificadas na mesma tabela.
 - Pelo menos um caminho de diretório de partição é um prefixo de pelo menos outro caminho de diretório de partição, por exemplo, `s3://bucket/table/p=a` é um prefixo de `s3://bucket/table/p=a b`.
 - O primeiro caractere que segue o prefixo no outro diretório de partição tem um valor UTF -8 menor que o / caractere (U+002F). Por exemplo, o caractere de espaço (U+0020) que ocorre entre a e b em `s3://bucket/table/p=a b` se enquadra nessa categoria. Observe que existem 14 outros caracteres que não são de controle: `!"#$%&'()*+,-.` Para obter mais informações, consulte a [tabela de codificação UTF -8 e caracteres Unicode](#).

Como solução alternativa para esse problema, defina a configuração `spark.sql.sources.fastS3PartitionDiscovery.enabled` como `false` na classificação `spark-defaults`.

- As conexões com EMR clusters da Amazon a partir do Amazon SageMaker Studio podem falhar intermitentemente com um código de resposta 403 Forbidden. Esse erro ocorre quando a configuração da IAM função no cluster leva mais de 60 segundos. Como solução alternativa, você pode instalar um EMR patch da Amazon para permitir novas tentativas e aumentar o tempo limite para um mínimo de 300 segundos. Use as etapas a seguir para aplicar a ação de bootstrap quando iniciar o cluster.

1. Baixe o script e os RPM arquivos de bootstrap do Amazon URIs S3 a seguir.

```
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. Carregue os arquivos da etapa anterior em um bucket do Amazon S3 de sua propriedade. O bucket deve estar no mesmo Região da AWS local em que você planeja iniciar o cluster.

- Inclua a seguinte ação de bootstrap ao iniciar seu EMR cluster. Substituir *bootstrap_URI* e *RPM_URI* com o correspondente URIs do Amazon S3.

```
--bootstrap-actions "Path=bootstrap_URI,Args=[RPM_URI]"
```

- Com as EMR versões 5.36.0 e 6.6.0 a 6.9.0 da Amazon, SecretAgent os componentes do RecordServer serviço podem sofrer perda de dados de log devido a uma configuração incorreta do padrão de nome de arquivo nas propriedades do Log4j2. A configuração incorreta faz com que os componentes gerem somente um arquivo de log por dia. Quando a estratégia de rotação ocorre, ela substitui o arquivo existente em vez de gerar um novo arquivo de log, conforme esperado. Como solução alternativa, use uma ação de bootstrap para gerar arquivos de log a cada hora e acrescentar um número inteiro de incremento automático no nome do arquivo para lidar com a rotação.

Para as versões EMR 6.6.0 a 6.9.0 da Amazon, use a seguinte ação de bootstrap ao iniciar um cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Para o Amazon EMR 5.36.0, use a seguinte ação de bootstrap ao iniciar um cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

- O Apache Flink fornece FileSystem conectores nativos S3 FileSystem e Hadoop, que permitem que os aplicativos criem FileSink e gravem os dados no Amazon S3. Isso FileSink falha com uma das duas exceções a seguir.

```
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:?]

```

Como solução alternativa, você pode instalar um EMR patch da Amazon, que corrige o problema acima no Flink. Para aplicar a ação de bootstrap quando iniciar o cluster, execute as etapas a seguir.

1. Baixe o flink-rpm no bucket Amazon S3. Seu RPM caminho é `s3://DOC-EXAMPLE-BUCKET/rpms/flink/`.
2. Faça o download do script e dos RPM arquivos de bootstrap do Amazon S3 usando o seguinte URI `regionName` Substitua pelo Região da AWS local em que você planeja iniciar o cluster.

```

s3://emr-data-access-control-regionName/customer-bootstrap-actions/gcsc/replace-
rpms.sh

```

3. O Hadoop 3.3.3 introduziu uma alteração em YARN ([YARN-9608](#)) que mantém os nós onde os contêineres eram executados em um estado de descomissionamento até que o aplicativo seja concluído. Essa alteração garante que dados locais, como dados embaralhados, não sejam perdidos e que você não precise executar o trabalho novamente. Na Amazon EMR 6.8.0 e 6.9.0, essa abordagem também pode levar à subutilização de recursos em clusters com ou sem escalabilidade gerenciada habilitada.

Com o [Amazon EMR 6.10.0](#), há uma solução alternativa para esse problema: definir o valor de `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications to in. false` `yarn-site.xml` Nas EMR versões 6.11.0 e superiores da Amazon, bem como nas 6.8.1, 6.9.1 e 6.10.1, a configuração é definida como padrão para resolver esse problema. `false`

Alterações, melhorias e problemas resolvidos

- Para a Amazon EMR versão 6.9.0 e posterior, todos os componentes instalados pela Amazon EMR que usam bibliotecas Log4j usam Log4j versão 2.17.1 ou posterior.
- Quando você usa o conector do DynamoDB com o Spark nas versões 6.6.0, 6.7.0 e 6.8.0 da EMR Amazon, todas as leituras da sua tabela retornam um resultado vazio, mesmo que a divisão de entrada faça referência a dados não vazios. A EMR versão 6.9.0 da Amazon corrige esse problema.

- O Amazon EMR 6.9.0 adiciona suporte limitado ao controle de acesso baseado em Lake Formation com o Apache Hudi ao ler dados usando o Spark. SQL O suporte é para SELECT consultas usando o Spark SQL e é limitado ao controle de acesso em nível de coluna. Para obter mais informações, consulte [Hudi e Lake Formation](#).
- Quando você usa o Amazon EMR 6.9.0 para criar um cluster Hadoop com [Node Labels](#) habilitados, as [YARNmétricas API](#) retornam informações agregadas em todas as partições, em vez da partição padrão. Para obter mais informações, consulte [YARN-11414](#).
- Com a EMR versão 6.9.0 da Amazon, atualizamos o Trino para a versão 398, que usa Java 17. A versão anterior compatível do Trino para Amazon EMR 6.8.0 era a Trino 388 em execução no Java 11. Para obter mais informações sobre essa alteração, consulte [Atualizações do Trino para Java 17](#) no blog do Trino.
- Esta versão corrige um problema de incompatibilidade de sequência de tempo entre o Apache BigTop e a Amazon EMR na sequência de inicialização EC2 do cluster. Essa incompatibilidade de sequência de tempo ocorre quando um sistema tenta realizar duas ou mais operações ao mesmo tempo em vez de fazê-las na sequência correta. Como resultado, determinadas configurações de cluster apresentaram tempos limite de inicialização da instância e tempos de inicialização do cluster mais lentos.
- Quando você inicia um cluster com a versão de patch mais recente do Amazon EMR 5.36 ou superior, 6.6 ou superior ou 7.0 ou superior, a Amazon EMR usa a versão mais recente do Amazon Linux 2023 ou Amazon Linux 2 para a Amazon padrão. EMR AMI Para obter mais informações, consulte [Usando o Amazon Linux padrão AMI para a Amazon EMR](#).

Note

Esta versão não recebe mais AMI atualizações automáticas, pois foi sucedida por mais 1 lançamento de patch. A versão de patch é indicada pelo número após o segundo ponto decimal (6.8.**1**). Para ver se você está usando a versão de patch mais recente, verifique as versões disponíveis no [Guia de lançamento](#) ou verifique o menu suspenso de EMRlançamentos da Amazon ao criar um cluster no console ou usar a ação [ListReleaseLabelsAPI](#) ou [list-release-labelsCLI](#). Para receber atualizações sobre novos lançamentos, assine o RSS feed do [What's new?](#) página.

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 709.1	4.14.348	23 de julho de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia), Ásia

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Pacífico (Hyderabad), Oriente Médio (), Europa (EspanhaUAE), Europa (Zurique), Ásia-Pacífico (Melbourne), Israel (Tel Aviv), Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 808.0	4.14.320	24 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), Israel (Tel Aviv).

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 727.0	4.14.320	14 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 719.0	4.14.320	2 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)
2.0.2023 628.0	4.14.318	12 de julho de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 612.0	4.14.314	23 de junho de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 504.1	4.14.313	16 de maio de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 418.0	4.14.311	3 de maio de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 404.1	4.14.311	18 de abril de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)
2.0.2023 404.0	4.14.311	10 de abril de 2023	Leste dos EUA (Norte da Virgínia), Europa (Paris)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 320.0	4.14.309	30 de março de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 307.0	4.14.305	15 de março de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 207.0	4.14.304	22 de fevereiro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 210.1	4.14.301	12 de janeiro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 103.3	4.14.296	5 de dezembro de 2022	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

Versões de componente 6.9.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são

a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado myapp-component com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como 2.2-amzn-2.

Componente	Version (Versão)	Descrição
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark SDK
delta	2.1.0	O Delta Lake é um formato de tabela aberto para grandes conjuntos de dados analíticos
emr-ddb	4.16.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	3.3.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.6.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-notebook-env	1.7.0	Ambiente Conda para bloco de anotações do emr, que inclui o jupyter enterprise gateway
emr-s3-dist-cp	2.23.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	2.2.0	EMRConector S3Select

Componente	Version (Versão)	Descrição
emrfs	2.54.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.15.2	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
flink-jobmanager-config	1.15.2	Gerenciando recursos em EMR nós para o Apache JobManager Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	3.3.3-amzn-1	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	3.3.3-amzn-1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	3.3.3-amzn-1	HDFScliente e biblioteca de linha de comando

Componente	Version (Versão)	Descrição
hadoop-hdfs-namenode	3.3.3-amzn-1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	3.3.3-amzn-1	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httpfs-server	3.3.3-amzn-1	HTTPendpoint para HDFS operações.
hadoop-kms-server	3.3.3-amzn-1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	3.3.3-amzn-1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	3.3.3-amzn-1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	3.3.3-amzn-1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	3.3.3-amzn-1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	2.4.13-amzn-0	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.

Componente	Version (Versão)	Descrição
hbase-region-server	2.4.13-amzn-0	Serviço para atender a uma ou mais HBase regiões.
hbase-client	2.4.13-amzn-0	HBase cliente de linha de comando.
hbase-rest-server	2.4.13-amzn-0	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	2.4.13-amzn-0	Serviço que fornece um endpoint Thrift para HBase
hbase-operator-tools	2.4.13-amzn-0	Ferramenta de reparo para HBase clusters Apache.
hcatalog-client	3.1.3-amzn-2	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	3.1.3-amzn-2	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	3.1.3-amzn-2	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	3.1.3-amzn-2	O cliente da linha de comando do Hive.
hive-hbase	3.1.3-amzn-2	Cliente do Hive-hbase.

Componente	Version (Versão)	Descrição
hive-metastore-server	3.1.3-amzn-2	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	3.1.3-amzn-2	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.12.1-amzn-0	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.
hudi-presto	0.12.1-amzn-0	Biblioteca de pacotes para executar o Presto com o Hudi.
hudi-trino	0.12.1-amzn-0	Biblioteca de pacotes para execução do Presto com o Hudi.
hudi-spark	0.12.1-amzn-0	Biblioteca de pacotes para executar o Spark com o Hudi.
hue-server	4.10.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
iceberg	0.14.1-amzn-0	Apache Iceberg é um formato de tabela aberta para conjuntos de dados analíticos imensos
jupyterhub	1.4.1	Servidor de vários usuários para blocos de anotações Jupyter

Componente	Version (Versão)	Descrição
livy-server	0.7.1-incubating	REST interface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mxnet	1.9.1	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mariadb-server	5.5.68+	Servidor de banco de dados MariaDB.
nvidia-cuda	11.7.0	Drivers NVIDIA e toolkit CUDA
oozie-client	5.2.1	O cliente da linha de comando do Oozie.
oozie-server	5.2.1	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	4.5.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	5.1.2	As bibliotecas do phoenix para servidor e cliente
phoenix-connectors	6.0.0- SNAPSHOT	Conectores do Apache Phoenix para Spark-3
phoenix-query-server	6.0.0	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API

Componente	Version (Versão)	Descrição
presto-coordinator	0.276-amzn-0	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.276-amzn-0	O serviço que executa partes de uma consulta.
presto-client	0.276-amzn-0	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
trino-coordinator	398-amzn-0	Serviço que aceita consultas e gerencia a execução de consultas entre trino-workers.
trino-worker	398-amzn-0	O serviço que executa partes de uma consulta.
trino-client	398-amzn-0	Cliente de linha de comando do Trino instalado em mestres de espera de um cluster de alta disponibilidade (HA) em que o servidor do Trino não é iniciado.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	4.0.2	O projeto R para computação estatística
ranger-kms-server	2.0.0	Sistema de gerenciamento de chaves Apache Ranger

Componente	Version (Versão)	Descrição
spark-client	3.3.0-amzn-1	Os clientes da linha de comando do Spark.
spark-history-server	3.3.0-amzn-1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	3.3.0-amzn-1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	3.3.0-amzn-1	Bibliotecas do Apache Spark necessárias para escravos. YARN
spark-rapids	22.08.0-amzn-0	RAPIDSPlugin Nvidia Spark que acelera o Apache Spark com o. GPUs
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	2.10.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.10.2-amzn-0	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.41+	HTTP Servidor Apache.
zeppelin-server	0.10.1	O notebook baseado na web que permite a análise de dados interativa.

Componente	Version (Versão)	Descrição
zookeeper-server	3.5.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.5.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 6.9.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Ocorrem ações de reconfiguração quando você especifica uma configuração para grupos de instâncias em um cluster em execução. A Amazon EMR só inicia ações de reconfiguração para as classificações que você modifica. Para ter mais informações, consulte [Reconfigurar um grupo de instâncias de um cluster em execução](#).

Classificações do emr-6.9.0

Classificações	Descrição	Ações de reconfiguração
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.	Restarts the Resource Manager service.
container-executor	Altere os valores no arquivo <code>container-executor.cfg</code> YARN do Hadoop.	Not available.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.	Not available.

Classificações	Descrição	Ações de reconfiguração
core-site	Alterar os valores no arquivo core-site.xml do 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	Alterar as configurações relacionadas ao docker.	Not available.
emrfs-site	Altere EMRFS as configurações.	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.

Classificações	Descrição	Ações de reconfiguração
flink-conf	Alterar as configurações do flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Alterar as configurações de log4j.properties no Flink.	Restarts Flink history server.
flink-log4j-session	Alterar as configurações de log4j-yarn-session.properties no Flink para sessão do Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.	Restarts Flink history server.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do 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.

Classificações	Descrição	Ações de reconfiguração
hadoop-log4j	Alterar os valores no arquivo log4j.properties do 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	Alterar a configuração do servidor SSL no Hadoop	Not available.
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop	Not available.
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Mude os valores no HBase ambiente.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.

Classificações	Descrição	Ações de reconfiguração
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.	Not available.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.	This classification should not be reconfigured.
hdfs-env	Mude os valores no HDFS ambiente.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Mude os valores no HCatalog ambiente.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.	Restarts Hive WebHCat server.

Classificações	Descrição	Ações de reconfiguração
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Configurações EMR selecionadas pela Amazon para o Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.	Not available.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.	Not available.
hive-env	Alterar os valores no ambiente do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log 4j2.properties do Hive.	Not available.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log 4j2.properties do Hive.	Not available.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.	Not available.

Classificações	Descrição	Ações de reconfiguração
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.	Not available.
hue-ini	Alterar os valores no arquivo ini do Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
https-env	Mude os valores no HTTPFS ambiente.	Restarts Hadoop Httpfs service.
https-site	Alterar os valores no arquivo https-site.xml do Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.	Not available.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.	Not available.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Altere os valores no ambiente do Hudi.	Not available.

Classificações	Descrição	Ações de reconfiguração
hudi-defaults	Alterar os valores no arquivo hudi-defaults.conf do Hudi.	Not available.
iceberg-defaults	Alterar os valores no arquivo iceberg-defaults.conf do Iceberg.	Not available.
delta-defaults	Alterar os valores no arquivo delta-defaults.conf do Delta.	Not available.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.	Not available.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.	Not available.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.	Not available.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.	Not available.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.	Restarts Livy Server.
livy-env	Alterar os valores no ambiente do Livy.	Restarts Livy Server.
livy-log4j2	Alterar as configurações de log4j2.properties do Livy.	Restarts Livy Server.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.

Classificações	Descrição	Ações de reconfiguração
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Alterar os valores no ambiente do Oozie.	Restarts Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.	Restarts Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.	Not available.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.	Not available.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.	Not available.
pig-env	Alterar os valores no ambiente do Pig.	Not available.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.	Restarts Oozie.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-log	Alterar os valores no arquivo log.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Alterar os valores no arquivo config.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.	Not available.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Altere os valores no arquivo node.properties do Presto.	Not available.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.	Not available.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.	Not available.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.	Not available.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.	Not available.
presto-connector-lakeformation	Alterar os valores no arquivo lakeformation.properties do Presto.	Restarts Presto-Server (for PrestoDB)

Classificações	Descrição	Ações de reconfiguração
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.	Not available.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.	Not available.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.	Not available.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.	Not available.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.	Not available.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.	Not available.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.	Not available.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.	Not available.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.	Not available.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.	Not available.
trino-log	Alterar os valores no arquivo log.properties do Trino.	Restarts Trino-Server (for Trino)

Classificações	Descrição	Ações de reconfiguração
trino-config	Alterar os valores no arquivo <code>config.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Alterar os valores no arquivo <code>password-authenticator.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-env	Alterar os valores no arquivo <code>trino-env.sh</code> do Trino.	Restarts Trino-Server (for Trino)
trino-node	Alterar os valores no arquivo <code>node.properties</code> do Trino.	Not available.
trino-connector-blackhole	Alterar os valores no arquivo <code>blackhole.properties</code> do Trino.	Not available.
trino-connector-cassandra	Alterar os valores no arquivo <code>cassandra.properties</code> do Trino.	Not available.
trino-connector-delta	Alterar os valores no arquivo <code>delta.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-connector-hive	Alterar os valores no arquivo <code>hive.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-exchange-manager	Alterar os valores no arquivo <code>exchange-manager.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Alterar os valores no arquivo <code>iceberg.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Alterar os valores no arquivo <code>jmx.properties</code> do Trino.	Not available.
trino-connector-kafka	Alterar os valores no arquivo <code>kafka.properties</code> do Trino.	Not available.

Classificações	Descrição	Ações de reconfiguração
trino-connector-localfile	Alterar os valores no arquivo localfile.properties do Trino.	Not available.
trino-connector-memory	Alterar os valores no arquivo memory.properties do Trino.	Not available.
trino-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Trino.	Not available.
trino-connector-mysql	Alterar os valores no arquivo mysql.properties do Trino.	Not available.
trino-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Trino.	Not available.
trino-connector-raptor	Alterar os valores no arquivo raptor.properties do Trino.	Not available.
trino-connector-redis	Alterar os valores no arquivo redis.properties do Trino.	Not available.
trino-connector-redshift	Alterar os valores no arquivo redshift.properties do Trino.	Not available.
trino-connector-tpch	Alterar os valores no arquivo tpch.properties do Trino.	Not available.
trino-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Trino.	Not available.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.	Restarts Ranger KMS Server.

Classificações	Descrição	Ações de reconfiguração
ranger-kms-env	Altere os valores no KMS ambiente Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Altere os valores no arquivo kms-log4j.properties do Ranger. KMS	Not available.
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.	Not available.
spark	Configurações EMR seleciona das pela Amazon para o Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Alterar os valores no ambiente do Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.	Not available.
spark-log4j2	Alterar os valores no arquivo log4j2.properties do Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Alterar os valores no ambiente do Sqoop.	Not available.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.	Not available.

Classificações	Descrição	Ações de reconfiguração
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.	Not available.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.	Restart Oozie and HiveServer2.
yarn-env	Mude os valores no YARN ambiente.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.	Restarts Zeppelin.
zeppelin-site	Alterar as definições da configuração em zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.	Restarts Zookeeper server.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.	Restarts Zookeeper server.

Log de alterações 6.9.0

Log de alterações da versão 6.9.0 e das notas da versão

Data	Evento	Descrição
2023-08-30	Atualizar notas da versão	Adicionada correção para um problema de incompatibilidade de sequência de tempo
2023-08-21	Atualizar notas da versão	Adicionado um problema conhecido com o Hadoop 3.3.3.
2023-07-26	Atualizar	Novos rótulos de versão 2.0.20230612.0 e 2.0.20230628.0 do sistema operacional.
2022-12-13	Notas da versão atualizadas	Recurso adicionado e problema conhecido para tempo de execução com SageMaker
2022-11-29	Notas da versão e documentação atualizadas	Adicionado atributo para a integração do Amazon Redshift para Apache Spark
2022-11-23	Notas da versão atualizadas	Removida entrada Log4j
2022-11-18	Implantação concluída	Amazon EMR 6.9 totalment e implantado em todas as regiões suportadas
2022-11-18	Publicação de documentos	Notas de lançamento do Amazon EMR 6.9 publicadas pela primeira vez

Data	Evento	Descrição
2022-11-14	Lançamento inicial	Amazon EMR 6.9 implantado em regiões comerciais limitadas

Amazon EMR versão 6.8.1

Versões 6.8.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-6.8.1	emr-6.8.0	emr-6.7.0	emr-6.6.0
AWS SDK para 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	2.12.15	2.12.15	2.12.15	2.12.10
AmazonCloudWatchAgent	-	-	-	-

	emr-6.8.1	emr-6.8.0	emr-6.7.0	emr-6.6.0
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
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

	emr-6.8.1	emr-6.8.0	emr-6.7.0	emr-6.6.0
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	0.10.1	0.10.1	0.10.0	0.10.0
ZooKeeper	3.5.10	3.5.10	3.5.7	3.5.7

Notas da versão 6.8.1

As notas de lançamento a seguir incluem informações sobre a EMR versão 6.8.1 da Amazon. As alterações são referentes à versão 6.8.0. Para obter informações sobre o cronograma da versão, consulte o [Log de alterações 6.8.1](#).

Alterações, melhorias e problemas resolvidos

- O Hadoop 3.3.3 introduziu uma alteração em YARN ([YARN-9608](#)) que mantém os nós onde os contêineres eram executados em um estado de descomissionamento até que o aplicativo seja concluído. Essa alteração garante que dados locais, como dados embaralhados, não sejam perdidos e que você não precise executar o trabalho novamente. Essa abordagem também pode levar à subutilização de recursos em clusters com ou sem o ajuste de escala gerenciado habilitado.

Com as EMR versões 6.11.0 e superiores da Amazon, bem como 6.8.1, 6.9.1 e 6.10.1, o valor de `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` é definido como `in` para resolver esse problema. `false yarn-site.xml`

Embora a correção resolva os problemas introduzidos pelo YARN -9608, ela pode fazer com que as tarefas do Hive falhem devido à perda aleatória de dados em clusters com escalabilidade gerenciada habilitada. Reduzimos esse risco nesta versão também ao configurar `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-shuffle-`

data para workloads do Hive. Essa configuração só está disponível nas EMR versões 6.11.0 e superiores da Amazon.

- O coletor de métricas não enviará métricas para o ambiente de gerenciamento após o failover do nó primário em clusters com a configuração de grupos de instâncias.
- Essa versão elimina novas tentativas de HTTP solicitações malsucedidas para endpoints do coletor de métricas.
- Esta versão inclui uma alteração que permite que clusters de alta disponibilidade se recuperem de um estado de falha após a reinicialização.
- Esta versão corrige um problema em que grandes exceções de estouro criadas por usuários causavam exceções de estouro.
- Esta versão corrige problemas de tempo limite com o processo de EMR reconfiguração da Amazon.
- Esta versão evita um problema em que a falha na reconfiguração pode interromper outros processos não relacionados.
- Esta versão inclui correções de segurança.
- Esta versão corrige um problema em que clusters que estão executando cargas de trabalho no Spark com a Amazon EMR podem receber silenciosamente resultados incorretos com `contains`, `startsWith`, e `endsWith like`. Esse problema ocorre quando você usa as expressões em campos particionados que têm metadados no Amazon EMR Hive3 Metastore Server (`hms`).
- Com o Amazon EMR 6.6.0 a 6.9.x, `INSERT` as consultas com partição dinâmica e uma cláusula `ORDER BY` ou `SORT BY` sempre terão dois redutores. Esse problema é causado pela OSS alteração [HIVE-20703](#), que coloca a otimização dinâmica da partição de classificação sob decisão baseada em custos. Se sua workload não exigir a classificação de partições dinâmicas, recomendamos que você defina a propriedade `hive.optimize.sort.dynamic.partition.threshold` como `-1` para desabilitar o novo atributo e obter o número de redutores calculado corretamente. Esse problema foi corrigido no OSS Hive como parte do [HIVE-22269](#) e foi corrigido no Amazon 6.10.0. EMR
- O Hive pode sofrer perda de dados quando você o usa HDFS como um diretório temporário e habilita a mesclagem de arquivos pequenos, e a tabela contém caminhos de partição estáticos.
- Esta versão corrige um problema de desempenho com o Hive se a mesclagem de arquivos pequenos (desativada por padrão) estiver ativada no final do ETL trabalho.
- Esta versão corrige um problema de limitação no lado do Glue quando não há funções definidas pelo usuário (`udf`).

- Esta versão corrige um problema que exclui registros de contêineres pelo serviço de agregação de registros de nós antes que o carregador de registros possa enviá-los para o S3 em caso de descomissionamento. YARN
- Esta versão corrige o manuseio de arquivos compactados/arquivados com rastreamento persistente de arquivos de armazenamento para. HBase
- Esta versão corrige um problema que afetou a performance do Spark quando você definiu um valor de `true` padrão para a configuração `spark.yarn.heterogeneousExecutors.enabled` no `spark-defaults.conf`.
- Esta versão corrige um problema com a falha do Reduce Task em ler dados embaralhados. O problema causou falhas na consulta do Hive com um erro de memória corrompida.
- Esta versão corrige um problema que fazia com que o provisionador de nós falhasse se o serviço HDFS NameNode (NN) ficasse preso no modo de segurança durante a substituição do nó.
- Esta versão adiciona um novo mecanismo de repetição ao fluxo de trabalho de escalabilidade de EMR clusters para clusters que executam o Presto ou o Trino. Essa melhoria reduz o risco de que o redimensionamento do cluster fique paralisado indefinidamente devido a uma única falha na operação de redimensionamento. Ela também aprimora a utilização dos clusters, porque seu cluster aumenta e reduz a escala verticalmente com mais rapidez.
- Esta versão melhora a lógica de redução do cluster para que seu cluster não tente reduzir os nós principais abaixo da configuração do fator de HDFS replicação do cluster. Isso se alinha aos seus requisitos de redundância de dados e reduz a probabilidade de uma operação de escalabilidade paralisar.
- O daemon de gerenciamento de logs foi atualizado para identificar todos os logs que estão em uso ativo com identificadores de arquivos abertos no armazenamento da instância local e nos processos associados. Essa atualização garante que a Amazon exclua EMR adequadamente os arquivos e recupere o espaço de armazenamento depois que os registros forem arquivados no Amazon S3.
- Esta versão inclui um aprimoramento do daemon de gerenciamento de logs que exclui diretórios de etapas vazios e não utilizados no sistema de arquivos de cluster local. Um número excessivamente grande de diretórios vazios pode degradar o desempenho dos EMR daemons da Amazon e resultar na sobreutilização do disco.
- Esta versão corrige um problema que pode ocorrer quando você cria um nó de borda ao replicar um dos nós primários de um cluster com vários nós primários. O nó de borda replicado pode causar atrasos nas operações de redução da escala verticalmente ou resultar em alta utilização de memória nos nós primários. Para obter mais informações sobre como criar um nó de borda para

se comunicar com seu EMR cluster, consulte [Edge Node Creator](#) no `aws-samples` repositório em GitHub.

- Esta versão melhora o processo de automação que a Amazon EMR usa para remontar EBS volumes da Amazon em uma instância após uma reinicialização.
- Esta versão corrige um problema que resultou em lacunas intermitentes nas métricas do Hadoop que a Amazon publica EMR na Amazon. CloudWatch
- Esta versão corrige um problema com EMR clusters em que uma atualização no arquivo de YARN configuração que contém a lista de exclusão de nós do cluster é interrompida devido à utilização excessiva do disco. A atualização incompleta impede futuras operações de redução da escala verticalmente do cluster. Esta versão garante que o cluster permaneça íntegro e que as operações de escalabilidade funcionem conforme esperado.
- Esta versão aprimora o daemon de gerenciamento de logs no cluster para monitorar pastas de log adicionais em seu cluster. EMR Essa melhoria minimiza os cenários de utilização excessiva do disco.
- Esta versão reinicia automaticamente o daemon de gerenciamento de logs no cluster quando ele é interrompido. Essa melhoria reduz o risco de os nós parecerem não íntegros devido à utilização excessiva do disco.
- Esta versão adiciona suporte ao arquivamento de logs no Amazon S3 durante a redução da escala do cluster verticalmente. Anteriormente, só era possível arquivar arquivos de log no Amazon S3 durante o encerramento do cluster. A nova capacidade garante que os arquivos de log gerados no cluster persistam no Amazon S3 mesmo após o encerramento do nó. Para obter mais informações, consulte [Configurar registro em log e depuração do cluster](#).
- Esta versão corrige um problema que ocorria quando o Amazon S3 URI para uma ação de bootstrap terminava com um número de porta, por exemplo: `a.b.c.d:4345` A Amazon os EMR estava analisando incorretamente URIs, então qualquer ação de bootstrap associada falharia.
- Esta versão corrige um problema de incompatibilidade de sequência de tempo entre o Apache BigTop e a Amazon EMR na sequência de inicialização EC2 do cluster. Essa incompatibilidade de sequência de tempo ocorre quando um sistema tenta realizar duas ou mais operações ao mesmo tempo em vez de fazê-las na sequência correta. Como resultado, determinadas configurações de cluster apresentaram tempos limite de inicialização da instância e tempos de inicialização do cluster mais lentos.
- Quando você inicia um cluster com a versão de patch mais recente do Amazon EMR 5.36 ou superior, 6.6 ou superior ou 7.0 ou superior, a Amazon EMR usa a versão mais recente do Amazon Linux 2023 ou Amazon Linux 2 para a Amazon padrão. EMR AMI Para obter mais informações, consulte [Usando o Amazon Linux padrão AMI para a Amazon EMR](#).

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 709.1	4.14.348	23 de julho de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia), Ásia

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Pacífico (Hyderabad), Oriente Médio (), Europa (EspanhaUAE), Europa (Zurique), Ásia-Pacífico (Melbourne), Israel (Tel Aviv),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 223.0	4.14.336	8 de março de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Canadá (Central), Israel (Tel Aviv), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 131.0	4.14.336	14 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Canadá (Central), Israel (Tel Aviv), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 124.0	4.14.336	7 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Canadá (Central), Israel (Tel Aviv), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 109.0	4.14.334	24 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Canadá (Central), Israel (Tel Aviv), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 218.0	4.14.330	2 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Canadá (Central), Israel (Tel Aviv), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 206.0	4.14.330	22 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Canadá (Central), Israel (Tel Aviv), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 116.0	4.14.328	11 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Canadá (Central), Israel (Tel Aviv), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 101.0	4.14.327	16 de novembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Canadá (Central), Israel (Tel Aviv), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 020.1	4.14.326	7 de novembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Canadá (Central), Israel (Tel Aviv), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 012.1	4.14.326	26 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Canadá (Central), Israel (Tel Aviv), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 926.0	4.14.322	19 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Canadá (Central), Israel (Tel Aviv), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 906.0	4.14.322	4 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Canadá (Central)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 822.0	4.14.322	30 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Canadá (Central)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 808.0	4.14.320	24 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Canadá (Central)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 727.0	4.14.320	14 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central)

Versões de componente 6.8.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são

a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado myapp-component com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como 2.2-amzn-2.

Componente	Version (Versão)	Descrição
aws-sagemaker-spark-sdk	1.4.2	Amazon SageMaker Spark SDK
emr-ddb	4.16.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	3.2.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.5.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-notebook-env	1.7.0	Ambiente Conda para bloco de anotações do emr, que inclui o jupyter enterprise gateway
emr-s3-dist-cp	2.22.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	2.1.0	EMRConector S3Select
emrfs	2.53.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
flink-client	1.15.1	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
flink-jobmanager-config	1.15.1	Gerenciando recursos em EMR nós para o Apache JobManager Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	3.2.1-amzn-8.1	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	3.2.1-amzn-8.1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	3.2.1-amzn-8.1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	3.2.1-amzn-8.1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.

Componente	Version (Versão)	Descrição
hadoop-hdfs-journalnode	3.2.1-amzn-8.1	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httfs-server	3.2.1-amzn-8.1	HTTPendpoint para HDFS operações.
hadoop-kms-server	3.2.1-amzn-8.1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	3.2.1-amzn-8.1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	3.2.1-amzn-8.1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	3.2.1-amzn-8.1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	3.2.1-amzn-8.1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	2.4.12-amzn-0.1	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	2.4.12-amzn-0.1	Serviço para atender a uma ou mais HBase regiões.

Componente	Version (Versão)	Descrição
hbase-client	2.4.12-amzn-0.1	HBase cliente de linha de comando.
hbase-rest-server	2.4.12-amzn-0.1	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	2.4.12-amzn-0.1	Serviço que fornece um endpoint Thrift para HBase.
hbase-operator-tools	2.4.12-amzn-0.1	Ferramenta de reparo para HBase clusters Apache.
hcatalog-client	3.1.3-amzn-1.1	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	3.1.3-amzn-1.1	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	3.1.3-amzn-1.1	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	3.1.3-amzn-1.1	O cliente da linha de comando do Hive.
hive-hbase	3.1.3-amzn-1.1	Cliente do Hive-hbase.
hive-metastore-server	3.1.3-amzn-1.1	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL

Componente	Version (Versão)	Descrição
hive-server2	3.1.3-amzn-1.1	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.11.1-amzn-0	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.
hudi-presto	0.11.1-amzn-0	Biblioteca de pacotes para executar o Presto com o Hudi.
hudi-trino	0.11.1-amzn-0	Biblioteca de pacotes para execução do Presto com o Hudi.
hudi-spark	0.11.1-amzn-0	Biblioteca de pacotes para executar o Spark com o Hudi.
hue-server	4.10.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
iceberg	0.14.0-amzn-0	Apache Iceberg é um formato de tabela aberta para conjuntos de dados analíticos imensos
jupyterhub	1.4.1	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.7.1-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso

Componente	Version (Versão)	Descrição
mxnet	1.9.1	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mariadb-server	5.5.68+	Servidor de banco de dados MariaDB.
nvidia-cuda	11.7.0	Drivers NVIDIA e toolkit CUDA
oozie-client	5.2.1	O cliente da linha de comando do Oozie.
oozie-server	5.2.1	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	4.5.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	5.1.2	As bibliotecas do phoenix para servidor e cliente
phoenix-connectors	5.1.2	Conectores do Apache Phoenix para Spark-3
phoenix-query-server	5.1.2	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.273.3-amzn-0	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.273.3-amzn-0	O serviço que executa partes de uma consulta.

Componente	Version (Versão)	Descrição
presto-client	0.273.3-amzn-0	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
trino-coordinator	388-amzn-0	Serviço que aceita consultas e gerencia a execução de consultas entre trino-workers.
trino-worker	388-amzn-0	O serviço que executa partes de uma consulta.
trino-client	388-amzn-0	Cliente de linha de comando do Trino instalado em mestres de espera de um cluster de alta disponibilidade (HA) em que o servidor do Trino não é iniciado.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	4.0.2	O projeto R para computação estatística
ranger-kms-server	2.0.0	Sistema de gerenciamento de chaves Apache Ranger
spark-client	3.3.0-amzn-0.1	Os clientes da linha de comando do Spark.

Componente	Version (Versão)	Descrição
spark-history-server	3.3.0-amzn-0.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	3.3.0-amzn-0.1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	3.3.0-amzn-0.1	Bibliotecas do Apache Spark necessárias para escravos. YARN
spark-rapids	22.06.0-amzn-0	RAPIDSPlugin Nvidia Spark que acelera o Apache Spark com o. GPUs
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	2.9.1	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.2	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.41+	HTTPServidor Apache.
zeppelin-server	0.10.1	O notebook baseado na web que permite a análise de dados interativa.

Componente	Version (Versão)	Descrição
zookeeper-server	3.5.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.5.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 6.8.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Ocorrem ações de reconfiguração quando você especifica uma configuração para grupos de instâncias em um cluster em execução. A Amazon EMR só inicia ações de reconfiguração para as classificações que você modifica. Para ter mais informações, consulte [Reconfigurar um grupo de instâncias de um cluster em execução](#).

Classificações do emr-6.8.1

Classificações	Descrição	Ações de reconfiguração
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.	Restarts the Resource Manager service.
container-executor	Altere os valores no arquivo <code>container-executor.cfg</code> YARN do Hadoop.	Not available.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.	Not available.

Classificações	Descrição	Ações de reconfiguração
core-site	Alterar os valores no arquivo core-site.xml do 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	Alterar as configurações relacionadas ao docker.	Not available.
emrfs-site	Altere EMRFS as configurações.	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.

Classificações	Descrição	Ações de reconfiguração
flink-conf	Alterar as configurações do flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Alterar as configurações de log4j.properties no Flink.	Restarts Flink history server.
flink-log4j-session	Alterar as configurações de log4j-yarn-session.properties no Flink para sessão do Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.	Restarts Flink history server.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do 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.

Classificações	Descrição	Ações de reconfiguração
hadoop-log4j	Alterar os valores no arquivo log4j.properties do 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	Alterar a configuração do servidor SSL no Hadoop	Not available.
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop	Not available.
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Mude os valores no HBase ambiente.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.

Classificações	Descrição	Ações de reconfiguração
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.	Not available.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.	This classification should not be reconfigured.
hdfs-env	Mude os valores no HDFS ambiente.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Mude os valores no HCatalog ambiente.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.	Restarts Hive WebHCat server.

Classificações	Descrição	Ações de reconfiguração
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Configurações EMR selecionadas pela Amazon para o Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.	Not available.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.	Not available.
hive-env	Alterar os valores no ambiente do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log 4j2.properties do Hive.	Not available.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log 4j2.properties do Hive.	Not available.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.	Not available.

Classificações	Descrição	Ações de reconfiguração
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.	Not available.
hue-ini	Alterar os valores no arquivo ini do Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
https-env	Mude os valores no HTTPFS ambiente.	Restarts Hadoop Httpfs service.
https-site	Alterar os valores no arquivo https-site.xml do Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.	Not available.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.	Not available.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Altere os valores no ambiente do Hudi.	Not available.

Classificações	Descrição	Ações de reconfiguração
hudi-defaults	Alterar os valores no arquivo hudi-defaults.conf do Hudi.	Not available.
iceberg-defaults	Alterar os valores no arquivo iceberg-defaults.conf do Iceberg.	Not available.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.	Not available.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.	Not available.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.	Not available.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.	Not available.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.	Restarts Livy Server.
livy-env	Alterar os valores no ambiente do Livy.	Restarts Livy Server.
livy-log4j	Alterar as configurações de log4j.properties no Livy.	Restarts Livy Server.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.

Classificações	Descrição	Ações de reconfiguração
oozie-env	Alterar os valores no ambiente do Oozie.	Restarts Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.	Restarts Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.	Not available.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.	Not available.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.	Not available.
pig-env	Alterar os valores no ambiente do Pig.	Not available.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.	Restarts Oozie.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.	Not available.
presto-log	Alterar os valores no arquivo log.properties do Presto.	Restarts Presto-Server (for PrestoDB)

Classificações	Descrição	Ações de reconfiguração
presto-config	Alterar os valores no arquivo <code>config.properties</code> do Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Altere os valores no arquivo <code>password-authenticator.properties</code> do Presto.	Not available.
presto-env	Altere os valores no arquivo <code>presto-env.sh</code> do Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Altere os valores no arquivo <code>node.properties</code> do Presto.	Not available.
presto-connector-blackhole	Alterar os valores no arquivo <code>blackhole.properties</code> do Presto.	Not available.
presto-connector-cassandra	Alterar os valores no arquivo <code>cassandra.properties</code> do Presto.	Not available.
presto-connector-hive	Alterar os valores no arquivo <code>hive.properties</code> do Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Alterar os valores no arquivo <code>jmx.properties</code> do Presto.	Not available.
presto-connector-kafka	Alterar os valores no arquivo <code>kafka.properties</code> do Presto.	Not available.
presto-connector-localfile	Alterar os valores no arquivo <code>localfile.properties</code> do Presto.	Not available.
presto-connector-memory	Alterar os valores no arquivo <code>memory.properties</code> do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-mongodb	Alterar os valores no arquivo <code>mongodb.properties</code> do Presto.	Not available.
presto-connector-mysql	Alterar os valores no arquivo <code>mysql.properties</code> do Presto.	Not available.
presto-connector-postgresql	Alterar os valores no arquivo <code>postgresql.properties</code> do Presto.	Not available.
presto-connector-raptor	Alterar os valores no arquivo <code>raptor.properties</code> do Presto.	Not available.
presto-connector-redis	Alterar os valores no arquivo <code>redis.properties</code> do Presto.	Not available.
presto-connector-redshift	Alterar os valores no arquivo de propriedades <code>redshift</code> do Presto.	Not available.
presto-connector-tpch	Alterar os valores no arquivo <code>tpch.properties</code> do Presto.	Not available.
presto-connector-tpcds	Alterar os valores no arquivo <code>tpcds.properties</code> do Presto.	Not available.
trino-log	Alterar os valores no arquivo <code>log.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-config	Alterar os valores no arquivo <code>config.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Alterar os valores no arquivo <code>password-authenticator.properties</code> do Trino.	Restarts Trino-Server (for Trino)

Classificações	Descrição	Ações de reconfiguração
trino-env	Alterar os valores no arquivo trino-env.sh do Trino.	Restarts Trino-Server (for Trino)
trino-node	Alterar os valores no arquivo node.properties do Trino.	Not available.
trino-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Trino.	Not available.
trino-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Trino.	Not available.
trino-connector-hive	Alterar os valores no arquivo hive.properties do Trino.	Restarts Trino-Server (for Trino)
trino-connector-iceberg	Alterar os valores no arquivo iceberg.properties do Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Alterar os valores no arquivo jmx.properties do Trino.	Not available.
trino-connector-kafka	Alterar os valores no arquivo kafka.properties do Trino.	Not available.
trino-connector-localfile	Alterar os valores no arquivo localfile.properties do Trino.	Not available.
trino-connector-memory	Alterar os valores no arquivo memory.properties do Trino.	Not available.
trino-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Trino.	Not available.
trino-connector-mysql	Alterar os valores no arquivo mysql.properties do Trino.	Not available.
trino-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Trino.	Not available.

Classificações	Descrição	Ações de reconfiguração
trino-connector-raptor	Alterar os valores no arquivo raptor.properties do Trino.	Not available.
trino-connector-redis	Alterar os valores no arquivo redis.properties do Trino.	Not available.
trino-connector-redshift	Alterar os valores no arquivo redshift.properties do Trino.	Not available.
trino-connector-tpch	Alterar os valores no arquivo tpch.properties do Trino.	Not available.
trino-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Trino.	Not available.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Altere os valores no arquivo kms-log4j.properties do Ranger. KMS	Not available.
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.	Not available.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.	This property modifies spark-defaults. See actions there.

Classificações	Descrição	Ações de reconfiguração
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Alterar os valores no ambiente do Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.	Not available.
spark-log4j2	Alterar os valores no arquivo log4j2.properties do Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Alterar os valores no ambiente do Sqoop.	Not available.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.	Not available.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.	Not available.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.	Restart Oozie and HiveServer2.
yarn-env	Mude os valores no YARN ambiente.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.

Classificações	Descrição	Ações de reconfiguração
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.	Restarts Zeppelin.
zeppelin-site	Alterar as definições da configuração em zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.	Restarts Zookeeper server.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.	Restarts Zookeeper server.

Log de alterações 6.8.1

Log de alterações da versão 6.8.1 e das notas da versão

Data	Evento	Descrição
2023-08-30	Atualizar notas da versão	Adicionadas várias correções relacionadas ao ambiente de gerenciamento às notas da versão
2023-08-21	Publicação de documentos	Notas de lançamento do Amazon EMR 6.8.1 publicadas pela primeira vez

Data	Evento	Descrição
2023-08-16	Implantação concluída	Amazon EMR 6.8.1 totalment e implantado em todas as regiões suportadas
2023-08-04	Lançamento inicial	Amazon EMR 6.8.1 foi implantado pela primeira vez em regiões comerciais limitadas

Amazon EMR versão 6.8.0

Versões 6.8.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-6.8.0	emr-6.7.0	emr-6.6.0	emr-6.5.0
AWS SDK para Java	1.12.170	1.12.170	1.12.170	1.12.31

	emr-6.8.0	emr-6.7.0	emr-6.6.0	emr-6.5.0
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12.15	2.12.15	2.12.10	2.12.10
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
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

	emr-6.8.0	emr-6.7.0	emr-6.6.0	emr-6.5.0
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	0.10.1	0.10.0	0.10.0	0.10.0
ZooKeeper	3.5.10	3.5.7	3.5.7	3.5.7

Notas da versão 6.8.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 6.8.0 da Amazon. As alterações são referentes à versão 6.7.0.

Novos atributos

- O recurso Amazon EMR Steps agora é compatível com endpoint e JDBC /clients Apache Livy. ODBC Para obter mais informações, consulte [Configurar funções de tempo de execução para EMR as etapas da Amazon](#).
- A EMR versão 6.8.0 da Amazon vem com a versão 2.4.12 do ApacheHBase. Com essa HBase versão, você pode arquivar e excluir suas HBase tabelas. O processamento de arquivos do Amazon S3 renomeia todos os arquivos da tabela para o diretório de arquivos. Isso pode ser um processo custoso e demorado. Agora, você pode pular o processamento de arquivos e rapidamente eliminar e excluir tabelas grandes. Para ter mais informações, consulte [Usando a HBase concha](#).

Problemas conhecidos

- O Hadoop 3.3.3 introduziu uma alteração em YARN ([YARN-9608](#)) que mantém os nós onde os contêineres eram executados em um estado de descomissionamento até que o aplicativo seja concluído. Essa alteração garante que dados locais, como dados embaralhados, não sejam perdidos e que você não precise executar o trabalho novamente. Na Amazon EMR 6.8.0 e 6.9.0, essa abordagem também pode levar à subutilização de recursos em clusters com ou sem escalabilidade gerenciada habilitada.

Com o [Amazon EMR 6.10.0](#), há uma solução alternativa para esse problema: definir o valor de `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` to `in.false` em `yarn-site.xml`. Nas EMR versões 6.11.0 e superiores da Amazon, bem como nas 6.8.1, 6.9.1 e 6.10.1, a configuração é definida como padrão para resolver esse problema. `false`

Alterações, melhorias e problemas resolvidos

- Quando a Amazon EMR lança 6.5.0, 6.6.0 ou 6.7.0 leu as tabelas do Apache Phoenix por meio do shell do Apache Spark, a Amazon produziu uma. `EMRNoSuchMethodError` A EMR versão 6.8.0 da Amazon corrige esse problema.
- A EMR versão 6.8.0 da Amazon vem com o [Apache Hudi](#) 0.11.1; no entanto, os clusters da EMR Amazon 6.8.0 também são compatíveis com o código aberto do Hudi 0.12.0. `hudi-spark3.3-bundle_2.12`
- A EMR versão 6.8.0 da Amazon vem com o Apache Spark 3.3.0. Esta versão do Spark usa o Apache Log4j 2 e o arquivo `log4j2.properties` para configurar o Log4j nos processos do Spark. Se você usa o Spark no cluster ou cria EMR clusters com parâmetros de configuração personalizados e deseja atualizar para a EMR versão 6.8.0 da Amazon, você deve migrar para a nova classificação de `spark-log4j2` configuração e formato de chave do Apache Log4j 2. Para ter mais informações, consulte [Migrar do Apache Log4j 1.x para Log4j 2.x](#).
- Quando você inicia um cluster com a versão de patch mais recente do Amazon EMR 5.36 ou superior, 6.6 ou superior ou 7.0 ou superior, a Amazon EMR usa a versão mais recente do Amazon Linux 2023 ou Amazon Linux 2 para a Amazon padrão. EMR AMI Para obter mais informações, consulte [Usando o Amazon Linux padrão AMI para a Amazon EMR](#).

Note

Esta versão não recebe mais AMI atualizações automáticas, pois foi sucedida por mais 1 lançamento de patch. A versão de patch é indicada pelo número após o segundo

ponto decimal (6.8.1). Para ver se você está usando a versão de patch mais recente, verifique as versões disponíveis no [Guia de lançamento](#) ou verifique o menu suspenso de EMRlançamentos da Amazon ao criar um cluster no console ou usar a ação [ListReleaseLabels](#)API ou [list-release-labels](#)CLI. Para receber atualizações sobre novos lançamentos, assine o RSS feed do [What's new?](#) página.

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 709.1	4.14.348	23 de julho de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia), Ásia

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Pacífico (Hyderabad), Oriente Médio (), Europa (EspanhaUAE), Europa (Zurique), Ásia-Pacífico (Melbourne), Israel (Tel Aviv)
2.0.2023 808.0	4.14.320	24 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 727.0	4.14.320	14 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 719.0	4.14.320	2 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(Bahrein), Oriente Médio (), Canadá (UAE Central)
2.0.2023 628.0	4.14.318	12 de julho de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 612.0	4.14.314	23 de junho de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 504.1	4.14.313	16 de maio de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 418.0	4.14.311	3 de maio de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 404.1	4.14.311	18 de abril de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)
2.0.2023 404.0	4.14.311	10 de abril de 2023	Leste dos EUA (Norte da Virgínia), Europa (Paris)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 320.0	4.14.309	30 de março de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 307.0	4.14.305	15 de março de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 207.0	4.14.304	22 de fevereiro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 119.1	4.14.301	3 de fevereiro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 210.1	4.14.301	22 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 103.3	4.14.296	5 de dezembro de 2022	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 004.0	4.14.294	2 de novembro de 2022	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 912.1	4.14.291	6 de setembro de 2022	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

Problemas conhecidos

- Quando você usa o conector do DynamoDB com o Spark nas versões 6.6.0, 6.7.0 e 6.8.0 da EMR Amazon, todas as leituras da sua tabela retornam um resultado vazio, mesmo que a divisão de entrada faça referência a dados não vazios. Isso ocorre porque o Spark 3.2.0 define `spark.hadoopRDD.ignoreEmptySplits` como `true` por padrão. Como solução alternativa,

defina explicitamente `spark.hadoopRDD.ignoreEmptySplits` como `false`. A EMR versão 6.9.0 da Amazon corrige esse problema.

- Quando você usa o Spark com a formatação de localização de partições do Hive para ler dados no Amazon S3 e executa o Spark nas EMR versões 5.30.0 a 5.36.0 e 6.2.0 a 6.9.0 da Amazon, você pode encontrar um problema que impede seu cluster de ler os dados corretamente. Isso poderá acontecer se suas partições tiverem todas as características a seguir:
 - Duas ou mais partições são verificadas na mesma tabela.
 - Pelo menos um caminho de diretório de partição é um prefixo de pelo menos outro caminho de diretório de partição, por exemplo, `s3://bucket/table/p=a` é um prefixo de `s3://bucket/table/p=a b`.
 - O primeiro caractere que segue o prefixo no outro diretório de partição tem um valor UTF -8 menor que o / caractere (U+002F). Por exemplo, o caractere de espaço (U+0020) que ocorre entre a e b em `s3://bucket/table/p=a b` se enquadra nessa categoria. Observe que existem 14 outros caracteres que não são de controle: `!"#$%&'()*+,-.` Para obter mais informações, consulte a [tabela de codificação UTF -8 e caracteres Unicode](#).

Como solução alternativa para esse problema, defina a configuração `spark.sql.sources.fastS3PartitionDiscovery.enabled` como `false` na classificação `spark-defaults`.

- Com as EMR versões 5.36.0 e 6.6.0 a 6.9.0 da Amazon, `SecretAgent` os componentes do `RecordServer` serviço podem sofrer perda de dados de log devido a uma configuração incorreta do padrão de nome de arquivo nas propriedades do `Log4j2`. A configuração incorreta faz com que os componentes gerem somente um arquivo de log por dia. Quando a estratégia de rotação ocorre, ela substitui o arquivo existente em vez de gerar um novo arquivo de log, conforme esperado. Como solução alternativa, use uma ação de bootstrap para gerar arquivos de log a cada hora e acrescentar um número inteiro de incremento automático no nome do arquivo para lidar com a rotação.

Para as versões EMR 6.6.0 a 6.9.0 da Amazon, use a seguinte ação de bootstrap ao iniciar um cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Para o Amazon EMR 5.36.0, use a seguinte ação de bootstrap ao iniciar um cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

Para obter mais informações sobre o cronograma da versão, consulte o [log de alterações](#).

Versões de componente 6.8.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion-amzn-EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.4.2	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.16.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	3.2.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.5.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-notebook-env</code>	1.7.0	Ambiente Conda para bloco de anotações do emr, que

Componente	Version (Versão)	Descrição
		inclui o jupyter enterprise gateway
emr-s3-dist-cp	2.22.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	2.1.0	EMRConector S3Select
emrfs	2.53.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.15.1	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
flink-jobmanager-config	1.15.1	Gerenciando recursos em EMR nós para o Apache JobManager Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.

Componente	Version (Versão)	Descrição
hadoop-client	3.2.1-amzn-8	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	3.2.1-amzn-8	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	3.2.1-amzn-8	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	3.2.1-amzn-8	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	3.2.1-amzn-8	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httpfs-server	3.2.1-amzn-8	HTTPendpoint para HDFS operações.
hadoop-kms-server	3.2.1-amzn-8	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	3.2.1-amzn-8	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	3.2.1-amzn-8	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	3.2.1-amzn-8	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.

Componente	Version (Versão)	Descrição
hadoop-yarn-timeline-server	3.2.1-amzn-8	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	2.4.12-amzn-0	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	2.4.12-amzn-0	Serviço para atender a uma ou mais HBase regiões.
hbase-client	2.4.12-amzn-0	HBase cliente de linha de comando.
hbase-rest-server	2.4.12-amzn-0	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	2.4.12-amzn-0	Serviço que fornece um endpoint Thrift para HBase.
hbase-operator-tools	2.4.12-amzn-0	Ferramenta de reparo para HBase clusters Apache.
hcatalog-client	3.1.3-amzn-1	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	3.1.3-amzn-1	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.

Componente	Version (Versão)	Descrição
hcatalog-webhcat-server	3.1.3-amzn-1	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	3.1.3-amzn-1	O cliente da linha de comando do Hive.
hive-hbase	3.1.3-amzn-1	Cliente do Hive-hbase.
hive-metastore-server	3.1.3-amzn-1	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	3.1.3-amzn-1	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.11.1-amzn-0	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.
hudi-presto	0.11.1-amzn-0	Biblioteca de pacotes para executar o Presto com o Hudi.
hudi-trino	0.11.1-amzn-0	Biblioteca de pacotes para execução do Presto com o Hudi.
hudi-spark	0.11.1-amzn-0	Biblioteca de pacotes para executar o Spark com o Hudi.
hue-server	4.10.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop

Componente	Version (Versão)	Descrição
iceberg	0.14.0-amzn-0	Apache Iceberg é um formato de tabela aberta para conjuntos de dados analíticos imensos
jupyterhub	1.4.1	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.7.1-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mxnet	1.9.1	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mariadb-server	5.5.68+	Servidor de banco de dados MariaDB.
nvidia-cuda	11.7.0	Drivers NVIDIA e toolkit CUDA
oozie-client	5.2.1	O cliente da linha de comando do Oozie.
oozie-server	5.2.1	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	4.5.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	5.1.2	As bibliotecas do phoenix para servidor e cliente

Componente	Version (Versão)	Descrição
phoenix-connectors	5.1.2	Conectores do Apache Phoenix para Spark-3
phoenix-query-server	5.1.2	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.273.3-amzn-0	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.273.3-amzn-0	O serviço que executa partes de uma consulta.
presto-client	0.273.3-amzn-0	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
trino-coordinator	388-amzn-0	Serviço que aceita consultas e gerencia a execução de consultas entre trino-workers.
trino-worker	388-amzn-0	O serviço que executa partes de uma consulta.
trino-client	388-amzn-0	Cliente de linha de comando do Trino instalado em mestres de espera de um cluster de alta disponibilidade (HA) em que o servidor do Trino não é iniciado.

Componente	Version (Versão)	Descrição
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	4.0.2	O projeto R para computação estatística
ranger-kms-server	2.0.0	Sistema de gerenciamento de chaves Apache Ranger
spark-client	3.3.0-amzn-0	Os clientes da linha de comando do Spark.
spark-history-server	3.3.0-amzn-0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	3.3.0-amzn-0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	3.3.0-amzn-0	Bibliotecas do Apache Spark necessárias para escravos. YARN
spark-rapids	22.06.0-amzn-0	RAPIDSPlugin Nvidia Spark que acelera o Apache Spark com o. GPUs
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	2.9.1	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.

Componente	Version (Versão)	Descrição
tez-on-yarn	0.9.2	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.41+	HTTPServidor Apache.
zeppelin-server	0.10.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.5.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.5.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 6.8.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Ocorrem ações de reconfiguração quando você especifica uma configuração para grupos de instâncias em um cluster em execução. A Amazon EMR só inicia ações de reconfiguração para as classificações que você modifica. Para ter mais informações, consulte [Reconfigurar um grupo de instâncias de um cluster em execução](#).

Classificações do emr-6.8.0

Classificações	Descrição	Ações de reconfiguração
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.	Restarts the Resource Manager service.

Classificações	Descrição	Ações de reconfiguração
container-executor	Altere os valores no arquivo container-executor.cfg YARN do Hadoop.	Not available.
container-log4j	Altere os valores no arquivo container-log4j.properties YARN do Hadoop.	Not available.
core-site	Alterar os valores no arquivo core-site.xml do 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	Alterar as configurações relacionadas ao docker.	Not available.

Classificações	Descrição	Ações de reconfiguração
emrfs-site	Altere EMRFS as configurações.	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	Alterar as configurações do <code>flink-conf.yaml</code> .	Restarts Flink history server.
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.	Restarts Flink history server.
flink-log4j-session	Alterar as configurações de <code>log4j-yarn-session.properties</code> no Flink para sessão do Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Alterar as configurações de <code>log4j-cli.properties</code> no Flink.	Restarts Flink history server.

Classificações	Descrição	Ações de reconfiguração
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do 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	Alterar os valores no arquivo log4j.properties do 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	Alterar a configuração do servidor SSL no Hadoop	Not available.
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop	Not available.

Classificações	Descrição	Ações de reconfiguração
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Mude os valores no HBase ambiente.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.	Not available.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.	This classification should not be reconfigured.
hdfs-env	Mude os valores no HDFS ambiente.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classificações	Descrição	Ações de reconfiguração
hdfs-site	Altere os valores no HDFS hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Mude os valores no HCatalog ambiente.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Configurações EMR selecionadas pela Amazon para o Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.	Not available.

Classificações	Descrição	Ações de reconfiguração
hive-parquet-logging	Alterar os valores no arquivo <code>parquet-logging.properties</code> do Hive.	Not available.
hive-env	Alterar os valores no ambiente do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive <code>schemaTool CLI</code> commands to verify <code>hive-metastore</code> .
hive-exec-log4j2	Altere os valores no arquivo <code>hive-exec-log 4j2.properties</code> do Hive.	Not available.
hive-llap-daemon-log4j2	Altere os valores no arquivo <code>llap-daemon-log 4j2.properties</code> do Hive.	Not available.
hive-log4j2	Alterar os valores no arquivo <code>hive-log4j2.properties</code> do Hive.	Not available.
hive-site	Alterar os valores no arquivo <code>hive-site.xml</code> do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive <code>schemaTool CLI</code> commands to verify <code>hive-metastore</code> . Also restarts Oozie and Zeppelin.
hiveserver2-site	Alterar os valores no arquivo <code>hiveserver2-site.xml</code> do Hive Server2.	Not available.
hue-ini	Alterar os valores no arquivo <code>ini</code> do Hue	Restarts Hue. Also activates Hue <code>config override CLI</code> commands to pick up new configurations.

Classificações	Descrição	Ações de reconfiguração
httpfs-env	Mude os valores no HTTPFS ambiente.	Restarts Hadoop Httpfs service.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.	Not available.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.	Not available.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Altere os valores no ambiente do Hudi.	Not available.
hudi-defaults	Alterar os valores no arquivo hudi-defaults.conf do Hudi.	Not available.
iceberg-defaults	Alterar os valores no arquivo iceberg-defaults.conf do Iceberg.	Not available.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.	Not available.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.	Not available.

Classificações	Descrição	Ações de reconfiguração
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.	Not available.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.	Not available.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.	Restarts Livy Server.
livy-env	Alterar os valores no ambiente do Livy.	Restarts Livy Server.
livy-log4j	Alterar as configurações de log4j.properties no Livy.	Restarts Livy Server.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Alterar os valores no ambiente do Oozie.	Restarts Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.	Restarts Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.	Not available.

Classificações	Descrição	Ações de reconfiguração
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.	Not available.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.	Not available.
pig-env	Alterar os valores no ambiente do Pig.	Not available.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.	Restarts Oozie.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.	Not available.
presto-log	Alterar os valores no arquivo log.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Alterar os valores no arquivo config.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.	Not available.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Altere os valores no arquivo node.properties do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.	Not available.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.	Not available.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.	Not available.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.	Not available.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.	Not available.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.	Not available.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.	Not available.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.	Not available.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.	Not available.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-redis	Alterar os valores no arquivo <code>redis.properties</code> do Presto.	Not available.
presto-connector-redshift	Alterar os valores no arquivo de propriedades <code>redshift</code> do Presto.	Not available.
presto-connector-tpch	Alterar os valores no arquivo <code>tpch.properties</code> do Presto.	Not available.
presto-connector-tpcds	Alterar os valores no arquivo <code>tpcds.properties</code> do Presto.	Not available.
trino-log	Alterar os valores no arquivo <code>log.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-config	Alterar os valores no arquivo <code>config.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Alterar os valores no arquivo <code>password-authenticator.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-env	Alterar os valores no arquivo <code>trino-env.sh</code> do Trino.	Restarts Trino-Server (for Trino)
trino-node	Alterar os valores no arquivo <code>node.properties</code> do Trino.	Not available.
trino-connector-blackhole	Alterar os valores no arquivo <code>blackhole.properties</code> do Trino.	Not available.
trino-connector-cassandra	Alterar os valores no arquivo <code>cassandra.properties</code> do Trino.	Not available.
trino-connector-hive	Alterar os valores no arquivo <code>hive.properties</code> do Trino.	Restarts Trino-Server (for Trino)

Classificações	Descrição	Ações de reconfiguração
trino-connector-iceberg	Alterar os valores no arquivo iceberg.properties do Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Alterar os valores no arquivo jmx.properties do Trino.	Not available.
trino-connector-kafka	Alterar os valores no arquivo kafka.properties do Trino.	Not available.
trino-connector-localfile	Alterar os valores no arquivo localfile.properties do Trino.	Not available.
trino-connector-memory	Alterar os valores no arquivo memory.properties do Trino.	Not available.
trino-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Trino.	Not available.
trino-connector-mysql	Alterar os valores no arquivo mysql.properties do Trino.	Not available.
trino-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Trino.	Not available.
trino-connector-raptor	Alterar os valores no arquivo raptor.properties do Trino.	Not available.
trino-connector-redis	Alterar os valores no arquivo redis.properties do Trino.	Not available.
trino-connector-redshift	Alterar os valores no arquivo redshift.properties do Trino.	Not available.
trino-connector-tpch	Alterar os valores no arquivo tpch.properties do Trino.	Not available.
trino-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Trino.	Not available.

Classificações	Descrição	Ações de reconfiguração
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Altere os valores no arquivo kms-log4j.properties do Ranger. KMS	Not available.
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.	Not available.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Alterar os valores no ambiente do Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.	Not available.
spark-log4j2	Alterar os valores no arquivo log4j2.properties do Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.	Restarts Spark history server and Spark thrift server.

Classificações	Descrição	Ações de reconfiguração
sqoop-env	Alterar os valores no ambiente do Sqoop.	Not available.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.	Not available.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.	Not available.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.	Restart Oozie and HiveServer2.
yarn-env	Mude os valores no YARN ambiente.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.	Restarts Zeppelin.
zeppelin-site	Alterar as definições da configuração em zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.	Restarts Zookeeper server.

Classificações	Descrição	Ações de reconfiguração
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.	Restarts Zookeeper server.

Log de alterações 6.8.0

Log de alterações da versão 6.8.0 e das notas da versão

Data	Evento	Descrição
2023-08-21	Atualizar	Adicionado um problema conhecido com o Hadoop 3.3.3.
2023-07-26	Atualizar	Novos rótulos de versão 2.0.20230612.0 e 2.0.20230628.0 do sistema operacional.
2022-09-06	Implantação concluída	Amazon EMR 6.8 totalmente e implantado em todas as regiões suportadas
2022-09-06	Publicação inicial	Notas de lançamento do Amazon EMR 6.8 publicadas pela primeira vez
2022-08-31	Lançamento inicial	Amazon EMR 6.8 lançado para regiões comerciais limitadas

Amazon EMR versão 6.7.0

Versões 6.7.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-6.7.0	emr-6.6.0	emr-6.5.0	emr-6.4.0
AWS SDK para 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	2.12.15	2.12.10	2.12.10	2.12.10
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

	emr-6.7.0	emr-6.6.0	emr-6.5.0	emr-6.4.0
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	-
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

	emr-6.7.0	emr-6.6.0	emr-6.5.0	emr-6.4.0
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

Notas da versão 6.7.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 6.7.0 da Amazon. As alterações são referentes à versão 6.6.0.

Data da versão inicial: 15 de julho de 2022

Novos atributos

- A Amazon EMR agora oferece suporte ao Apache Spark 3.2.1, Apache Hive 3.1.3, HUDI 0.11, PrestoDB 0.272 e Trino 0.378.
- Suporta controles de acesso baseados em IAM Role e Lake Formation com EMR etapas (Spark, Hive) para Amazon em clusters. EMR EC2
- Ele é compatível com instruções de definição de dados do Apache Spark em clusters habilitados para Apache Ranger. Isso agora inclui suporte para aplicações do Trino lendo e gravando metadados do Apache Hive em clusters habilitados para Apache Ranger. Para obter mais informações, consulte [Habilitar a governança federada usando Trino e Apache Ranger na Amazon. EMR](#)
- Quando você inicia um cluster com a versão de patch mais recente do Amazon EMR 5.36 ou superior, 6.6 ou superior ou 7.0 ou superior, a Amazon EMR usa a versão mais recente do Amazon Linux 2023 ou Amazon Linux 2 para a Amazon padrão. EMR AMI Para obter mais informações, consulte [Usando o Amazon Linux padrão AMI para a Amazon EMR](#).

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 709.1	4.14.348	23 de julho de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia), Ásia

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Pacífico (Hyderabad), Oriente Médio (UAE), Europa (Espanha), Europa (Zurique)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 223.0	4.14.336	8 de março de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 131.0	4.14.336	14 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 124.0	4.14.336	7 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 109.0	4.14.334	24 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 218.0	4.14.330	2 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 206.0	4.14.330	22 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 116.0	4.14.328	11 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 101.0	4.14.327	16 de novembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 020.1	4.14.326	7 de novembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 012.1	4.14.326	26 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 926.0	4.14.322	19 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 906.0	4.14.322	4 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 822.0	4.14.322	30 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 808.0	4.14.320	24 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 727.0	4.14.320	14 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 719.0	4.14.320	2 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Oriente Médio (), Canadá (Central) UAE

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 628.0	4.14.318	12 de julho de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 612.0	4.14.314	23 de junho de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 504.1	4.14.313	16 de maio de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 418.0	4.14.311	3 de maio de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 404.1	4.14.311	18 de abril de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)
2.0.2023 404.0	4.14.311	10 de abril de 2023	Leste dos EUA (Norte da Virgínia), Europa (Paris)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 320.0	4.14.309	30 de março de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 307.0	4.14.305	15 de março de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 207.0	4.14.304	22 de fevereiro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 119.1	4.14.301	3 de fevereiro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 210.1	4.14.301	22 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 103.3	4.14.296	5 de dezembro de 2022	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 004.0	4.14.294	2 de novembro de 2022	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 912.1	4.14.291	7 de outubro de 2022	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)
2.0.2022 719.0	4.14.287	10 de agosto de 2022	us-west-1 , eu-west-3 , eu-north-1 , ap-south-1 , me-south-1

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 606.1	4.14.281	15 de julho de 2022	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

Problemas conhecidos

- Quando as EMR versões 6.5.0, 6.6.0 ou 6.7.0 da Amazon leem as tabelas do Apache Phoenix por meio do shell do Apache Spark, a ocorre porque a Amazon usa um `NoSuchMethodError` código incorreto. `EMR Hbase.compat.version` A EMR versão 6.8.0 da Amazon corrige esse problema.

- Quando você usa o conector do DynamoDB com o Spark nas versões 6.6.0, 6.7.0 e 6.8.0 da EMR Amazon, todas as leituras da sua tabela retornam um resultado vazio, mesmo que a divisão de entrada faça referência a dados não vazios. Isso ocorre porque o Spark 3.2.0 define `spark.hadoopRDD.ignoreEmptySplits` como `true` por padrão. Como solução alternativa, defina explicitamente `spark.hadoopRDD.ignoreEmptySplits` como `false`. A EMR versão 6.9.0 da Amazon corrige esse problema.
- Quando você usa o Spark com a formatação de localização de partições do Hive para ler dados no Amazon S3 e executa o Spark nas EMR versões 5.30.0 a 5.36.0 e 6.2.0 a 6.9.0 da Amazon, você pode encontrar um problema que impede seu cluster de ler os dados corretamente. Isso poderá acontecer se suas partições tiverem todas as características a seguir:
 - Duas ou mais partições são verificadas na mesma tabela.
 - Pelo menos um caminho de diretório de partição é um prefixo de pelo menos outro caminho de diretório de partição, por exemplo, `s3://bucket/table/p=a` é um prefixo de `s3://bucket/table/p=a b`.
 - O primeiro caractere que segue o prefixo no outro diretório de partição tem um valor UTF -8 menor que o / caractere (U+002F). Por exemplo, o caractere de espaço (U+0020) que ocorre entre a e b em `s3://bucket/table/p=a b` se enquadra nessa categoria. Observe que existem 14 outros caracteres que não são de controle: `!"#$%&'()*+,-.` Para obter mais informações, consulte a [tabela de codificação UTF -8 e caracteres Unicode](#).

Como solução alternativa para esse problema, defina a configuração `spark.sql.sources.fastS3PartitionDiscovery.enabled` como `false` na classificação `spark-defaults`.

- Com as EMR versões 5.36.0 e 6.6.0 a 6.9.0 da Amazon, `SecretAgent` os componentes do `RecordServer` serviço podem sofrer perda de dados de log devido a uma configuração incorreta do padrão de nome de arquivo nas propriedades do `Log4j2`. A configuração incorreta faz com que os componentes gerem somente um arquivo de log por dia. Quando a estratégia de rotação ocorre, ela substitui o arquivo existente em vez de gerar um novo arquivo de log, conforme esperado. Como solução alternativa, use uma ação de bootstrap para gerar arquivos de log a cada hora e acrescentar um número inteiro de incremento automático no nome do arquivo para lidar com a rotação.

Para as versões EMR 6.6.0 a 6.9.0 da Amazon, use a seguinte ação de bootstrap ao iniciar um cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Para o Amazon EMR 5.36.0, use a seguinte ação de bootstrap ao iniciar um cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

- O `GetClusterSessionCredentials` API não é compatível com clusters executados na Amazon EMR 6.7 ou inferior.

Versões de componente 6.7.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.16.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	3.2.0	Bibliotecas convenientes para o ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
emr-kinesis	3.5.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-notebook-env	1.6.0	Ambiente Conda para bloco de anotações do emr, que inclui o jupyter enterprise gateway
emr-s3-dist-cp	2.22.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	2.1.0	EMRConector S3Select
emrfs	2.52.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.14.2	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
flink-jobmanager-config	1.14.2	Gerenciando recursos em EMR nós para o Apache JobManager Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.

Componente	Version (Versão)	Descrição
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	3.2.1-amzn-7	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	3.2.1-amzn-7	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	3.2.1-amzn-7	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	3.2.1-amzn-7	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	3.2.1-amzn-7	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httpfs-server	3.2.1-amzn-7	HTTPendpoint para HDFS operações.
hadoop-kms-server	3.2.1-amzn-7	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	3.2.1-amzn-7	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.

Componente	Version (Versão)	Descrição
hadoop-yarn-nodemanager	3.2.1-amzn-7	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	3.2.1-amzn-7	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	3.2.1-amzn-7	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	2.4.4-amzn-3	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	2.4.4-amzn-3	Serviço para atender a uma ou mais HBase regiões.
hbase-client	2.4.4-amzn-3	HBasecliente de linha de comando.
hbase-rest-server	2.4.4-amzn-3	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	2.4.4-amzn-3	Serviço que fornece um endpoint Thrift para. HBase
hbase-operator-tools	2.4.4-amzn-3	Ferramenta de reparo para HBase clusters Apache.
hcatalog-client	3.1.3-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.

Componente	Version (Versão)	Descrição
hcatalog-server	3.1.3-amzn-0	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	3.1.3-amzn-0	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	3.1.3-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	3.1.3-amzn-0	Cliente do Hive-hbase.
hive-metastore-server	3.1.3-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	3.1.3-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.11.0-amzn-0	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.
hudi-presto	0.11.0-amzn-0	Biblioteca de pacotes para executar o Presto com o Hudi.
hudi-trino	0.11.0-amzn-0	Biblioteca de pacotes para execução do Presto com o Hudi.

Componente	Version (Versão)	Descrição
hudi-spark	0.11.0-amzn-0	Biblioteca de pacotes para executar o Spark com o Hudi.
hue-server	4.10.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
iceberg	0.13.1-amzn-0	Apache Iceberg é um formato de tabela aberta para conjuntos de dados analíticos imensos
jupyterhub	1.4.1	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.7.1-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mxnet	1.8.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mariadb-server	5.5.68+	Servidor de banco de dados MariaDB.
nvidia-cuda	11.0.194	Drivers NVIDIA e toolkit CUDA
oozie-client	5.2.1	O cliente da linha de comando do Oozie.
oozie-server	5.2.1	O serviço que aceita solicitações de fluxo de trabalho do Oozie.

Componente	Version (Versão)	Descrição
opencv	4.5.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	5.1.2	As bibliotecas do phoenix para servidor e cliente
phoenix-connectors	5.1.2	Conectores do Apache Phoenix para Spark-3
phoenix-query-server	5.1.2	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.272-amzn-0	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.272-amzn-0	O serviço que executa partes de uma consulta.
presto-client	0.272-amzn-0	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
trino-coordinator	378-amzn-0	Serviço que aceita consultas e gerencia a execução de consultas entre trino-workers.
trino-worker	378-amzn-0	O serviço que executa partes de uma consulta.

Componente	Version (Versão)	Descrição
trino-client	378-amzn-0	Cliente de linha de comando do Trino instalado em mestres de espera de um cluster de alta disponibilidade (HA) em que o servidor do Trino não é iniciado.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	4.0.2	O projeto R para computação estatística
ranger-kms-server	2.0.0	Sistema de gerenciamento de chaves Apache Ranger
spark-client	3.2.1-amzn-0	Os clientes da linha de comando do Spark.
spark-history-server	3.2.1-amzn-0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	3.2.1-amzn-0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	3.2.1-amzn-0	Bibliotecas do Apache Spark necessárias para escravos. YARN
spark-rapids	22.02.0-amzn-1	RAPIDSPlugin Nvidia Spark que acelera o Apache Spark com o. GPUs

Componente	Version (Versão)	Descrição
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	2.4.1	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.2	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.41+	HTTPServidor Apache.
zeppelin-server	0.10.0	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.5.7	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.5.7	ZooKeeper cliente de linha de comando.

Classificações de configuração 6.7.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Ocorrem ações de reconfiguração quando você especifica uma configuração para grupos de instâncias em um cluster em execução. A Amazon EMR só inicia ações de reconfiguração para as

classificações que você modifica. Para ter mais informações, consulte [Reconfigurar um grupo de instâncias de um cluster em execução](#).

Classificações do emr-6.7.0

Classificações	Descrição	Ações de reconfiguração
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.	Restarts the ResourceManager service.
container-executor	Altere os valores no arquivo container-executor.cfg YARN do Hadoop.	Not available.
container-log4j	Altere os valores no arquivo container-log4j.properties YARN do Hadoop.	Not available.
core-site	Alterar os valores no arquivo core-site.xml do 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	Alterar as configurações relacionadas ao docker.	Not available.
emrfs-site	Altere EMRFS as configurações.	Restarts the Hadoop HDFS services Namenode,

Classificações	Descrição	Ações de reconfiguração
		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	Alterar as configurações do <code>flink-conf.yaml</code> .	Restarts Flink history server.
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.	Restarts Flink history server.
flink-log4j-session	Alterar as configurações de <code>log4j-yarn-session.properties</code> no Flink para sessão do Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Alterar as configurações de <code>log4j-cli.properties</code> no Flink.	Restarts Flink history server.

Classificações	Descrição	Ações de reconfiguração
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do 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	Alterar os valores no arquivo log4j.properties do 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	Alterar a configuração do servidor SSL no Hadoop	Not available.
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop	Not available.

Classificações	Descrição	Ações de reconfiguração
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Mude os valores no HBase ambiente.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.	Not available.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.	This classification should not be reconfigured.
hdfs-env	Mude os valores no HDFS ambiente.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classificações	Descrição	Ações de reconfiguração
hdfs-site	Altere os valores no HDFS hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Mude os valores no HCatalog ambiente.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Configurações EMR selecionadas pela Amazon para o Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.	Not available.

Classificações	Descrição	Ações de reconfiguração
hive-parquet-logging	Alterar os valores no arquivo <code>parquet-logging.properties</code> do Hive.	Not available.
hive-env	Alterar os valores no ambiente do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Altere os valores no arquivo <code>hive-exec-log 4j2.properties</code> do Hive.	Not available.
hive-llap-daemon-log4j2	Altere os valores no arquivo <code>llap-daemon-log 4j2.properties</code> do Hive.	Not available.
hive-log4j2	Alterar os valores no arquivo <code>hive-log4j2.properties</code> do Hive.	Not available.
hive-site	Alterar os valores no arquivo <code>hive-site.xml</code> do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Alterar os valores no arquivo <code>hiveserver2-site.xml</code> do Hive Server2.	Not available.
hue-ini	Alterar os valores no arquivo <code>ini</code> do Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.

Classificações	Descrição	Ações de reconfiguração
httpfs-env	Mude os valores no HTTPFS ambiente.	Restarts Hadoop Httpfs service.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.	Not available.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.	Not available.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Altere os valores no ambiente do Hudi.	Not available.
hudi-defaults	Alterar os valores no arquivo hudi-defaults.conf do Hudi.	Not available.
iceberg-defaults	Alterar os valores no arquivo iceberg-defaults.conf do Iceberg.	Not available.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.	Not available.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.	Not available.

Classificações	Descrição	Ações de reconfiguração
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.	Not available.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.	Not available.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.	Restarts Livy Server.
livy-env	Alterar os valores no ambiente do Livy.	Restarts Livy Server.
livy-log4j	Alterar as configurações de log4j.properties no Livy.	Restarts Livy Server.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Alterar os valores no ambiente do Oozie.	Restarts Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.	Restarts Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.	Not available.

Classificações	Descrição	Ações de reconfiguração
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.	Not available.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.	Not available.
pig-env	Alterar os valores no ambiente do Pig.	Not available.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.	Restarts Oozie.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.	Not available.
presto-log	Alterar os valores no arquivo log.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Alterar os valores no arquivo config.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.	Not available.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Altere os valores no arquivo node.properties do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.	Not available.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.	Not available.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.	Not available.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.	Not available.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.	Not available.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.	Not available.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.	Not available.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.	Not available.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.	Not available.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-redis	Alterar os valores no arquivo <code>redis.properties</code> do Presto.	Not available.
presto-connector-redshift	Alterar os valores no arquivo de propriedades <code>redshift</code> do Presto.	Not available.
presto-connector-tpch	Alterar os valores no arquivo <code>tpch.properties</code> do Presto.	Not available.
presto-connector-tpcds	Alterar os valores no arquivo <code>tpcds.properties</code> do Presto.	Not available.
trino-log	Alterar os valores no arquivo <code>log.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-config	Alterar os valores no arquivo <code>config.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Alterar os valores no arquivo <code>password-authenticator.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-env	Alterar os valores no arquivo <code>trino-env.sh</code> do Trino.	Restarts Trino-Server (for Trino)
trino-node	Alterar os valores no arquivo <code>node.properties</code> do Trino.	Not available.
trino-connector-blackhole	Alterar os valores no arquivo <code>blackhole.properties</code> do Trino.	Not available.
trino-connector-cassandra	Alterar os valores no arquivo <code>cassandra.properties</code> do Trino.	Not available.
trino-connector-hive	Alterar os valores no arquivo <code>hive.properties</code> do Trino.	Restarts Trino-Server (for Trino)

Classificações	Descrição	Ações de reconfiguração
trino-connector-iceberg	Alterar os valores no arquivo iceberg.properties do Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Alterar os valores no arquivo jmx.properties do Trino.	Not available.
trino-connector-kafka	Alterar os valores no arquivo kafka.properties do Trino.	Not available.
trino-connector-localfile	Alterar os valores no arquivo localfile.properties do Trino.	Not available.
trino-connector-memory	Alterar os valores no arquivo memory.properties do Trino.	Not available.
trino-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Trino.	Not available.
trino-connector-mysql	Alterar os valores no arquivo mysql.properties do Trino.	Not available.
trino-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Trino.	Not available.
trino-connector-raptor	Alterar os valores no arquivo raptor.properties do Trino.	Not available.
trino-connector-redis	Alterar os valores no arquivo redis.properties do Trino.	Not available.
trino-connector-redshift	Alterar os valores no arquivo redshift.properties do Trino.	Not available.
trino-connector-tpch	Alterar os valores no arquivo tpch.properties do Trino.	Not available.
trino-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Trino.	Not available.

Classificações	Descrição	Ações de reconfiguração
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Altere os valores no arquivo kms-log4j.properties do Ranger. KMS	Not available.
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.	Not available.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Alterar os valores no ambiente do Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.	Not available.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.	Restarts Spark history server and Spark thrift server.

Classificações	Descrição	Ações de reconfiguração
sqoop-env	Alterar os valores no ambiente do Sqoop.	Not available.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.	Not available.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.	Not available.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.	Restart Oozie and HiveServer2.
yarn-env	Mude os valores no YARN ambiente.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.	Restarts Zeppelin.
zeppelin-site	Alterar as definições da configuração em zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.	Restarts Zookeeper server.

Classificações	Descrição	Ações de reconfiguração
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.	Restarts Zookeeper server.

Amazon EMR versão 6.6.0

Versões 6.6.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-6.6.0	emr-6.5.0	emr-6.4.0	emr-6.3.1
AWS SDK para Java	1.12.170	1.12.31	1.12.31	1.11.977
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12.10	2.12.10	2.12.10	2.12.10
AmazonCloudWatchAgent	-	-	-	-

	emr-6.6.0	emr-6.5.0	emr-6.4.0	emr-6.3.1
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
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

	emr-6.6.0	emr-6.5.0	emr-6.4.0	emr-6.3.1
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

Notas da versão 6.6.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 6.6.0 da Amazon. As alterações são referentes à versão 6.5.0.

Data da versão inicial: 9 de maio de 2022

Data da documentação atualizada: 15 de junho de 2022

Novos atributos

- O Amazon EMR 6.6 agora é compatível com Apache Spark 3.2, Apache Spark RAPIDS 22.02, CUDA 11, Apache Hudi 0.10.1, Apache Iceberg 0.13, Trino 0.367 e PrestoDB 0.267.
- Quando você inicia um cluster com a versão de patch mais recente do Amazon EMR 5.36 ou superior, 6.6 ou superior ou 7.0 ou superior, a Amazon EMR usa a versão mais recente do Amazon Linux 2023 ou Amazon Linux 2 para a Amazon padrão. EMR AMI Para obter mais informações, consulte [Usando o Amazon Linux padrão AMI para a Amazon EMR](#).

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 709.1	4.14.348	23 de julho de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 223.0	4.14.336	8 de março de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 131.0	4.14.336	14 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 124.0	4.14.336	7 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 109.0	4.14.334	24 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 218.0	4.14.330	2 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 206.0	4.14.330	22 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 116.0	4.14.328	11 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 101.0	4.14.327	16 de novembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 020.1	4.14.326	7 de novembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 012.1	4.14.326	26 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 926.0	4.14.322	19 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 906.0	4.14.322	4 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 822.0	4.14.322	30 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 808.0	4.14.320	24 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 727.0	4.14.320	14 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 719.0	4.14.320	2 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Oriente Médio (), Canadá (Central) UAE

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 628.0	4.14.318	12 de julho de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 612.0	4.14.314	23 de junho de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 504.1	4.14.313	16 de maio de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 418.0	4.14.311	3 de maio de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 404.1	4.14.311	18 de abril de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)
2.0.2023 404.0	4.14.311	10 de abril de 2023	Leste dos EUA (Norte da Virgínia), Europa (Paris)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 320.0	4.14.309	30 de março de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 307.0	4.14.305	15 de março de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 207.0	4.14.304	22 de fevereiro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 119.1	4.14.301	3 de fevereiro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 210.1	4.14.301	22 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 103.3	4.14.296	5 de dezembro de 2022	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 004.0	4.14.294	2 de novembro de 2022	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 912.1	4.14.291	7 de outubro de 2022	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)
2.0.2022 805.0	4.14.287	30 de agosto de 2022	us-west-1

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 719.0	4.14.287	10 de agosto de 2022	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 426.0	4.14.281	10 de junho de 2022	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 406.1	4.14.275	2 de maio de 2022	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

- [Com o Amazon EMR 6.6 e versões posteriores, os aplicativos que usam o Log4j 1.x e o Log4j 2.x são atualizados para usar o Log4j 1.2.17 \(ou superior\) e o Log4j 2.17.1 \(ou superior\), respectivamente, e não exigem o uso das ações de bootstrap fornecidas para mitigar os problemas.](#) CVE
- [Escalabilidade gerenciada] Otimização de escalabilidade gerenciada de dados do Spark shuffle
 - Para EMR as versões 5.34.0 e posteriores da Amazon e as EMR versões 6.4.0 e posteriores, a escalabilidade gerenciada agora reconhece os dados do Spark shuffle (dados que o Spark

redistribui entre partições para realizar operações específicas). Para obter mais informações sobre operações aleatórias, consulte Como [usar a escalabilidade EMR gerenciada na Amazon no Guia de EMR Gerenciamento EMR da Amazon](#) e no Guia de Programação do [Spark](#).

- A partir do Amazon EMR 5.32.0 e 6.5.0, o dimensionamento dinâmico do executor para o Apache Spark está habilitado por padrão. Para ativar ou desativar esse atributo, você pode usar o parâmetro de configuração `spark.yarn.heterogeneousExecutors.enabled`.

Alterações, melhorias e problemas resolvidos

- A Amazon EMR reduz o tempo de inicialização do cluster em até 80 segundos, em média, para clusters que usam a AMI opção EMR padrão e instalam apenas aplicativos comuns, como Apache Hadoop, Apache Spark e Apache Hive.

Problemas conhecidos

- Quando as EMR versões 6.5.0, 6.6.0 ou 6.7.0 da Amazon leem as tabelas do Apache Phoenix por meio do shell do Apache Spark, a ocorre porque a Amazon usa um `NoSuchMethodError` código incorreto. EMR `Hbase.compat.version` A EMR versão 6.8.0 da Amazon corrige esse problema.
- Quando você usa o conector do DynamoDB com o Spark nas versões 6.6.0, 6.7.0 e 6.8.0 da EMR Amazon, todas as leituras da sua tabela retornam um resultado vazio, mesmo que a divisão de entrada faça referência a dados não vazios. Isso ocorre porque o Spark 3.2.0 define `spark.hadoopRDD.ignoreEmptySplits` como `true` por padrão. Como solução alternativa, defina explicitamente `spark.hadoopRDD.ignoreEmptySplits` como `false`. A EMR versão 6.9.0 da Amazon corrige esse problema.
- Em clusters de longa duração do Trino, o Amazon EMR 6.6.0 habilita os parâmetros de registro da coleta de lixo no `jvm.config` do Trino para obter melhores insights dos registros da coleta de lixo. Essa alteração anexa muitos logs da coleta de resíduos ao arquivo `launcher.log` (`/var/log/trino/launcher.log`). Se você estiver executando clusters Trino no Amazon EMR 6.6.0, poderá encontrar nós sem espaço em disco depois que o cluster estiver em execução por alguns dias devido aos registros anexados.

A solução alternativa para esse problema é executar o script abaixo como uma ação de bootstrap para desativar os parâmetros de registro da coleta de lixo no `jvm.config` ao criar ou clonar o cluster para o 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"
```

- Quando você usa o Spark com a formatação de localização de partições do Hive para ler dados no Amazon S3 e executa o Spark nas EMR versões 5.30.0 a 5.36.0 e 6.2.0 a 6.9.0 da Amazon, você pode encontrar um problema que impede seu cluster de ler os dados corretamente. Isso poderá acontecer se suas partições tiverem todas as características a seguir:
 - Duas ou mais partições são verificadas na mesma tabela.
 - Pelo menos um caminho de diretório de partição é um prefixo de pelo menos outro caminho de diretório de partição, por exemplo, s3://bucket/table/p=a é um prefixo de s3://bucket/table/p=a b.
 - O primeiro caractere que segue o prefixo no outro diretório de partição tem um valor UTF -8 menor que o / caractere (U+002F). Por exemplo, o caractere de espaço (U+0020) que ocorre entre a e b em s3://bucket/table/p=a b se enquadra nessa categoria. Observe que existem 14 outros caracteres que não são de controle: !"#\$%&'()*+,-. Para obter mais informações, consulte a [tabela de codificação UTF -8 e caracteres Unicode](#).

Como solução alternativa para esse problema, defina a configuração

`spark.sql.sources.fastS3PartitionDiscovery.enabled` como `false` na classificação `spark-defaults`.

- Com as EMR versões 5.36.0 e 6.6.0 a 6.9.0 da Amazon, `SecretAgent` os componentes do `RecordServer` serviço podem sofrer perda de dados de log devido a uma configuração incorreta do padrão de nome de arquivo nas propriedades do `Log4j2`. A configuração incorreta faz com que os componentes gerem somente um arquivo de log por dia. Quando a estratégia de rotação ocorre, ela substitui o arquivo existente em vez de gerar um novo arquivo de log, conforme esperado. Como solução alternativa, use uma ação de bootstrap para gerar arquivos de log a cada hora e acrescentar um número inteiro de incremento automático no nome do arquivo para lidar com a rotação.

Para as versões EMR 6.6.0 a 6.9.0 da Amazon, use a seguinte ação de bootstrap ao iniciar um cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Para o Amazon EMR 5.36.0, use a seguinte ação de bootstrap ao iniciar um cluster.


```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

Versões de componente 6.6.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion-amzn-EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.16.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	3.2.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.5.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-notebook-env</code>	1.5.0	Ambiente Conda para bloco de anotações do emr, que

Componente	Version (Versão)	Descrição
		inclui o jupyter enterprise gateway
emr-s3-dist-cp	2.20.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	2.1.0	EMRConector S3Select
emrfs	2.50.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.14.2	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
flink-jobmanager-config	1.14.2	Gerenciando recursos em EMR nós para o Apache JobManager Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.

Componente	Version (Versão)	Descrição
hadoop-client	3.2.1-amzn-6	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	3.2.1-amzn-6	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	3.2.1-amzn-6	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	3.2.1-amzn-6	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	3.2.1-amzn-6	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httpfs-server	3.2.1-amzn-6	HTTPendpoint para HDFS operações.
hadoop-kms-server	3.2.1-amzn-6	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	3.2.1-amzn-6	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	3.2.1-amzn-6	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	3.2.1-amzn-6	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.

Componente	Version (Versão)	Descrição
hadoop-yarn-timeline-server	3.2.1-amzn-6	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	2.4.4-amzn-2	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	2.4.4-amzn-2	Serviço para atender a uma ou mais HBase regiões.
hbase-client	2.4.4-amzn-2	HBase cliente de linha de comando.
hbase-rest-server	2.4.4-amzn-2	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	2.4.4-amzn-2	Serviço que fornece um endpoint Thrift para HBase.
hbase-operator-tools	2.4.4-amzn-2	Ferramenta de reparo para HBase clusters Apache.
hcatalog-client	3.1.2-amzn-7	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	3.1.2-amzn-7	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.

Componente	Version (Versão)	Descrição
hcatalog-webhcat-server	3.1.2-amzn-7	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	3.1.2-amzn-7	O cliente da linha de comando do Hive.
hive-hbase	3.1.2-amzn-7	Cliente do Hive-hbase.
hive-metastore-server	3.1.2-amzn-7	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	3.1.2-amzn-7	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.10.1-amzn-0	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.
hudi-presto	0.10.1-amzn-0	Biblioteca de pacotes para executar o Presto com o Hudi.
hudi-trino	0.10.1-amzn-0	Biblioteca de pacotes para execução do Presto com o Hudi.
hudi-spark	0.10.1-amzn-0	Biblioteca de pacotes para executar o Spark com o Hudi.
hue-server	4.10.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop

Componente	Version (Versão)	Descrição
iceberg	0.13.1	Apache Iceberg é um formato de tabela aberta para conjuntos de dados analíticos imensos
jupyterhub	1.4.1	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.7.1-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mxnet	1.8.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mariadb-server	5.5.68+	Servidor de banco de dados MariaDB.
nvidia-cuda	11.0.194	Drivers NVIDIA e toolkit CUDA
oozie-client	5.2.1	O cliente da linha de comando do Oozie.
oozie-server	5.2.1	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	4.5.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	5.1.2	As bibliotecas do phoenix para servidor e cliente

Componente	Version (Versão)	Descrição
phoenix-connectors	5.1.2	Conectores do Apache Phoenix para Spark-3
phoenix-query-server	5.1.2	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.267-amzn-0	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.267-amzn-0	O serviço que executa partes de uma consulta.
presto-client	0.267-amzn-0	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
trino-coordinator	367-amzn-0	Serviço que aceita consultas e gerencia a execução de consultas entre trino-workers.
trino-worker	367-amzn-0	O serviço que executa partes de uma consulta.
trino-client	367-amzn-0	Cliente de linha de comando do Trino instalado em mestres de espera de um cluster de alta disponibilidade (HA) em que o servidor do Trino não é iniciado.

Componente	Version (Versão)	Descrição
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	4.0.2	O projeto R para computação estatística
ranger-kms-server	2.0.0	Sistema de gerenciamento de chaves Apache Ranger
spark-client	3.2.0-amzn-0	Os clientes da linha de comando do Spark.
spark-history-server	3.2.0-amzn-0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	3.2.0-amzn-0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	3.2.0-amzn-0	Bibliotecas do Apache Spark necessárias para escravos. YARN
spark-rapids	22.02.0-amzn-0	RAPIDSPlugin Nvidia Spark que acelera o Apache Spark com o. GPUs
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	2.4.1	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.

Componente	Version (Versão)	Descrição
tez-on-yarn	0.9.2	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.41+	HTTPServidor Apache.
zeppelin-server	0.10.0	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.5.7	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.5.7	ZooKeeper cliente de linha de comando.

Classificações de configuração 6.6.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Ocorrem ações de reconfiguração quando você especifica uma configuração para grupos de instâncias em um cluster em execução. A Amazon EMR só inicia ações de reconfiguração para as classificações que você modifica. Para ter mais informações, consulte [Reconfigurar um grupo de instâncias de um cluster em execução](#).

Classificações do emr-6.6.0

Classificações	Descrição	Ações de reconfiguração
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.	Restarts the Resource Manager service.

Classificações	Descrição	Ações de reconfiguração
container-executor	Altere os valores no arquivo container-executor.cfg YARN do Hadoop.	Not available.
container-log4j	Altere os valores no arquivo container-log4j.properties YARN do Hadoop.	Not available.
core-site	Alterar os valores no arquivo core-site.xml do 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	Alterar as configurações relacionadas ao docker.	Not available.

Classificações	Descrição	Ações de reconfiguração
emrfs-site	Altere EMRFS as configurações.	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	Alterar as configurações do <code>flink-conf.yaml</code> .	Restarts Flink history server.
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.	Restarts Flink history server.
flink-log4j-session	Alterar as configurações de <code>log4j-yarn-session.properties</code> no Flink para sessão do Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Alterar as configurações de <code>log4j-cli.properties</code> no Flink.	Restarts Flink history server.

Classificações	Descrição	Ações de reconfiguração
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do 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	Alterar os valores no arquivo log4j.properties do 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	Alterar a configuração do servidor SSL no Hadoop	Not available.
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop	Not available.

Classificações	Descrição	Ações de reconfiguração
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Mude os valores no HBase ambiente.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.	Not available.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.	This classification should not be reconfigured.
hdfs-env	Mude os valores no HDFS ambiente.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classificações	Descrição	Ações de reconfiguração
hdfs-site	Altere os valores no HDFS hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Mude os valores no HCatalog ambiente.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Configurações EMR selecionadas pela Amazon para o Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.	Not available.

Classificações	Descrição	Ações de reconfiguração
hive-parquet-logging	Alterar os valores no arquivo <code>parquet-logging.properties</code> do Hive.	Not available.
hive-env	Alterar os valores no ambiente do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Altere os valores no arquivo <code>hive-exec-log 4j2.properties</code> do Hive.	Not available.
hive-llap-daemon-log4j2	Altere os valores no arquivo <code>llap-daemon-log 4j2.properties</code> do Hive.	Not available.
hive-log4j2	Alterar os valores no arquivo <code>hive-log4j2.properties</code> do Hive.	Not available.
hive-site	Alterar os valores no arquivo <code>hive-site.xml</code> do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Alterar os valores no arquivo <code>hiveserver2-site.xml</code> do Hive Server2.	Not available.
hue-ini	Alterar os valores no arquivo <code>ini</code> do Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.

Classificações	Descrição	Ações de reconfiguração
httpfs-env	Mude os valores no HTTPFS ambiente.	Restarts Hadoop Httpfs service.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.	Not available.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.	Not available.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Altere os valores no ambiente do Hudi.	Not available.
hudi-defaults	Alterar os valores no arquivo hudi-defaults.conf do Hudi.	Not available.
iceberg-defaults	Alterar os valores no arquivo iceberg-defaults.conf do Iceberg.	Not available.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.	Not available.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.	Not available.

Classificações	Descrição	Ações de reconfiguração
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.	Not available.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.	Not available.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.	Restarts Livy Server.
livy-env	Alterar os valores no ambiente do Livy.	Restarts Livy Server.
livy-log4j	Alterar as configurações de log4j.properties no Livy.	Restarts Livy Server.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Alterar os valores no ambiente do Oozie.	Restarts Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.	Restarts Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.	Not available.

Classificações	Descrição	Ações de reconfiguração
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.	Not available.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.	Not available.
pig-env	Alterar os valores no ambiente do Pig.	Not available.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.	Restarts Oozie.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.	Not available.
presto-log	Alterar os valores no arquivo log.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Alterar os valores no arquivo config.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.	Not available.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Altere os valores no arquivo node.properties do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.	Not available.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.	Not available.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.	Not available.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.	Not available.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.	Not available.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.	Not available.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.	Not available.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.	Not available.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.	Not available.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-redis	Alterar os valores no arquivo <code>redis.properties</code> do Presto.	Not available.
presto-connector-redshift	Alterar os valores no arquivo de propriedades <code>redshift</code> do Presto.	Not available.
presto-connector-tpch	Alterar os valores no arquivo <code>tpch.properties</code> do Presto.	Not available.
presto-connector-tpcds	Alterar os valores no arquivo <code>tpcds.properties</code> do Presto.	Not available.
trino-log	Alterar os valores no arquivo <code>log.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-config	Alterar os valores no arquivo <code>config.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Alterar os valores no arquivo <code>password-authenticator.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-env	Alterar os valores no arquivo <code>trino-env.sh</code> do Trino.	Restarts Trino-Server (for Trino)
trino-node	Alterar os valores no arquivo <code>node.properties</code> do Trino.	Not available.
trino-connector-blackhole	Alterar os valores no arquivo <code>blackhole.properties</code> do Trino.	Not available.
trino-connector-cassandra	Alterar os valores no arquivo <code>cassandra.properties</code> do Trino.	Not available.
trino-connector-hive	Alterar os valores no arquivo <code>hive.properties</code> do Trino.	Restarts Trino-Server (for Trino)

Classificações	Descrição	Ações de reconfiguração
trino-connector-iceberg	Alterar os valores no arquivo iceberg.properties do Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Alterar os valores no arquivo jmx.properties do Trino.	Not available.
trino-connector-kafka	Alterar os valores no arquivo kafka.properties do Trino.	Not available.
trino-connector-localfile	Alterar os valores no arquivo localfile.properties do Trino.	Not available.
trino-connector-memory	Alterar os valores no arquivo memory.properties do Trino.	Not available.
trino-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Trino.	Not available.
trino-connector-mysql	Alterar os valores no arquivo mysql.properties do Trino.	Not available.
trino-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Trino.	Not available.
trino-connector-raptor	Alterar os valores no arquivo raptor.properties do Trino.	Not available.
trino-connector-redis	Alterar os valores no arquivo redis.properties do Trino.	Not available.
trino-connector-redshift	Alterar os valores no arquivo redshift.properties do Trino.	Not available.
trino-connector-tpch	Alterar os valores no arquivo tpch.properties do Trino.	Not available.
trino-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Trino.	Not available.

Classificações	Descrição	Ações de reconfiguração
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Altere os valores no arquivo kms-log4j.properties do Ranger. KMS	Not available.
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.	Not available.
spark	Configurações EMR seleciona das pela Amazon para o Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Alterar os valores no ambiente do Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.	Not available.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.	Restarts Spark history server and Spark thrift server.

Classificações	Descrição	Ações de reconfiguração
sqoop-env	Alterar os valores no ambiente do Sqoop.	Not available.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.	Not available.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.	Not available.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.	Restart Oozie and HiveServer2.
yarn-env	Mude os valores no YARN ambiente.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.	Restarts Zeppelin.
zeppelin-site	Alterar as definições da configuração em zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.	Restarts Zookeeper server.

Classificações	Descrição	Ações de reconfiguração
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.	Restarts Zookeeper server.

Amazon EMR versão 6.5.0

Versões 6.5.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-6.5.0	emr-6.4.0	emr-6.3.1	emr-6.3.0
AWS SDK para Java	1.12.31	1.12.31	1.11.977	1.11.977
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12.10	2.12.10	2.12.10	2.12.10
AmazonCloudWatchAgent	-	-	-	-

	emr-6.5.0	emr-6.4.0	emr-6.3.1	emr-6.3.0
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
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

	emr-6.5.0	emr-6.4.0	emr-6.3.1	emr-6.3.0
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

Notas da versão 6.5.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 6.5.0 da Amazon. As alterações são referentes à versão 6.4.0.

Data da versão inicial: 20 de janeiro de 2022

Data da versão atualizada: 21 de março de 2022

Novos atributos

- [Escalabilidade gerenciada] Otimização de escalabilidade gerenciada de dados do Spark shuffle - Para EMR as versões 5.34.0 e posteriores da Amazon e as EMR versões 6.4.0 e posteriores, a escalabilidade gerenciada agora reconhece os dados do Spark shuffle (dados que o Spark redistribui entre partições para realizar operações específicas). Para obter mais informações sobre operações aleatórias, consulte Como [usar a escalabilidade EMR gerenciada na Amazon no Guia de EMR Gerenciamento EMR da Amazon](#) e no Guia de Programação do [Spark](#).
- A partir do Amazon EMR 5.32.0 e 6.5.0, o dimensionamento dinâmico do executor para o Apache Spark está habilitado por padrão. Para ativar ou desativar esse atributo, você pode usar o parâmetro de configuração `spark.yarn.heterogeneousExecutors.enabled`.
- Suporte para o formato de tabela aberta Apache Iceberg para conjuntos de dados analíticos imensos.
- Support para ranger-trino-plugin 2.0.1-amzn-1

- Suporte para toree 0.5.0

Alterações, melhorias e problemas resolvidos

- A versão de lançamento do Amazon EMR 6.5 agora oferece suporte ao Apache Iceberg 0.12.0 e fornece melhorias no tempo de execução com o Amazon Runtime para Apache Spark, o Amazon EMR Runtime para Presto e o Amazon Runtime para Apache Hive. EMR
- O [Apache Iceberg](#) é um formato de tabela aberta para grandes conjuntos de dados no Amazon S3 e fornece desempenho rápido de consultas em tabelas grandes, confirmações atômicas, gravações simultâneas e evolução de tabelas compatíveis. SQL Com a EMR versão 6.5, você pode usar o Apache Spark 3.1.2 com o formato de tabela Iceberg.
- O Apache Hudi 0.9 adiciona Spark e suporte SQLDDL. DML Isso permite que você crie e altere tabelas Hudi usando apenas declarações. SQL O Apache Hudi 0.9 também inclui melhorias na performance do lado da consulta e do lado do gravador.
- O Amazon EMR Runtime para Apache Hive melhora o desempenho do Apache Hive no Amazon S3 ao remover as operações de renomeação durante as operações de preparação e melhora o desempenho dos comandos metastore check () usados para reparar tabelas. MSCK

Problemas conhecidos

- Quando as EMR versões 6.5.0, 6.6.0 ou 6.7.0 da Amazon leem as tabelas do Apache Phoenix por meio do shell do Apache Spark, a ocorre porque a Amazon usa um NoSuchMethodError código incorreto. EMR Hbase.compat.version A EMR versão 6.8.0 da Amazon corrige esse problema.
- Os clusters do pacote do Hbase em alta disponibilidade (HA) apresentam falha no provisionamento com o tamanho de volume e o tipo de instância padrão. A solução alternativa para esse problema é aumentar o tamanho do volume raiz.
- Para usar as ações do Spark com o Apache Oozie, você deve adicionar a seguinte configuração ao seu arquivo workflow.xml do Oozie. Caso contrário, várias bibliotecas críticas, como a Hadoop e a Hadoop, EMRFS estarão ausentes do classpath dos executores do Spark que o Oozie lança.

```
<spark-opts>--conf spark.yarn.populateHadoopClasspath=true</spark-opts>
```

- Quando você usa o Spark com a formatação de localização de partições do Hive para ler dados no Amazon S3 e executa o Spark nas EMR versões 5.30.0 a 5.36.0 e 6.2.0 a 6.9.0 da Amazon, você

pode encontrar um problema que impede seu cluster de ler os dados corretamente. Isso poderá acontecer se suas partições tiverem todas as características a seguir:

- Duas ou mais partições são verificadas na mesma tabela.
- Pelo menos um caminho de diretório de partição é um prefixo de pelo menos outro caminho de diretório de partição, por exemplo, `s3://bucket/table/p=a` é um prefixo de `s3://bucket/table/p=a b`.
- O primeiro caractere que segue o prefixo no outro diretório de partição tem um valor UTF -8 menor que o / caractere (U+002F). Por exemplo, o caractere de espaço (U+0020) que ocorre entre a e b em `s3://bucket/table/p=a b` se enquadra nessa categoria. Observe que existem 14 outros caracteres que não são de controle: `!"#$%&'()*+,-.` Para obter mais informações, consulte a [tabela de codificação UTF -8 e caracteres Unicode](#).

Como solução alternativa para esse problema, defina a configuração `spark.sql.sources.fastS3PartitionDiscovery.enabled` como `false` na classificação `spark-defaults`.

Versões de componente 6.5.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark SDK

Componente	Version (Versão)	Descrição
emr-ddb	4.16.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	3.2.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.5.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-notebook-env	1.4.0	Ambiente Conda para bloco de anotações do emr, que inclui o jupyter enterprise gateway
emr-s3-dist-cp	2.19.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	2.1.0	EMRConector S3Select
emrfs	2.48.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.14.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
flink-jobmanager-config	1.14.0	Gerenciando recursos em EMR nós para o Apache JobManager Flink.

Componente	Version (Versão)	Descrição
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	3.2.1-amzn-5	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	3.2.1-amzn-5	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	3.2.1-amzn-5	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	3.2.1-amzn-5	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	3.2.1-amzn-5	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httpfs-server	3.2.1-amzn-5	HTTPendpoint para HDFS operações.

Componente	Version (Versão)	Descrição
hadoop-kms-server	3.2.1-amzn-5	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	3.2.1-amzn-5	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	3.2.1-amzn-5	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	3.2.1-amzn-5	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	3.2.1-amzn-5	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	2.4.4-amzn-1	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	2.4.4-amzn-1	Serviço para atender a uma ou mais HBase regiões.
hbase-client	2.4.4-amzn-1	HBasecliente de linha de comando.
hbase-rest-server	2.4.4-amzn-1	Serviço que fornece um RESTful HTTP endpoint paraHBase.

Componente	Version (Versão)	Descrição
hbase-thrift-server	2.4.4-amzn-1	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	3.1.2-amzn-6	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	3.1.2-amzn-6	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	3.1.2-amzn-6	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	3.1.2-amzn-6	O cliente da linha de comando do Hive.
hive-hbase	3.1.2-amzn-6	Cliente do Hive-hbase.
hive-metastore-server	3.1.2-amzn-6	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	3.1.2-amzn-6	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.9.0-amzn-1	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.

Componente	Version (Versão)	Descrição
hudi-presto	0.9.0-amzn-1	Biblioteca de pacotes para executar o Presto com o Hudi.
hudi-trino	0.9.0-amzn-1	Biblioteca de pacotes para execução do Presto com o Hudi.
hudi-spark	0.9.0-amzn-1	Biblioteca de pacotes para executar o Spark com o Hudi.
hue-server	4.9.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
iceberg	0.12.0	Apache Iceberg é um formato de tabela aberta para conjuntos de dados analíticos imensos
jupyterhub	1.4.1	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.7.1-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mxnet	1.8.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mariadb-server	5.5.68+	Servidor de banco de dados MariaDB.
nvidia-cuda	10.1.243	Drivers NVIDIA e toolkit CUDA

Componente	Version (Versão)	Descrição
oozie-client	5.2.1	O cliente da linha de comando do Oozie.
oozie-server	5.2.1	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	4.5.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	5.1.2	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	5.1.2	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.261-amzn-0	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.261-amzn-0	O serviço que executa partes de uma consulta.
presto-client	0.261-amzn-0	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
trino-coordinator	360	Serviço que aceita consultas e gerencia a execução de consultas entre trino-workers.

Componente	Version (Versão)	Descrição
trino-worker	360	O serviço que executa partes de uma consulta.
trino-client	360	Cliente de linha de comando do Trino instalado em mestres de espera de um cluster de alta disponibilidade (HA) em que o servidor do Trino não é iniciado.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	4.0.2	O projeto R para computação estatística
ranger-kms-server	2.0.0	Sistema de gerenciamento de chaves Apache Ranger
spark-client	3.1.2-amzn-1	Os clientes da linha de comando do Spark.
spark-history-server	3.1.2-amzn-1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	3.1.2-amzn-1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	3.1.2-amzn-1	Bibliotecas do Apache Spark necessárias para escravos. YARN

Componente	Version (Versão)	Descrição
spark-rapids	0.4.1	RAPIDSPlugin Nvidia Spark que acelera o Apache Spark com o. GPUs
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	2.4.1	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.2	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.41+	HTTPServidor Apache.
zeppelin-server	0.10.0	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.5.7	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.5.7	ZooKeeper cliente de linha de comando.

Classificações de configuração 6.5.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Ocorrem ações de reconfiguração quando você especifica uma configuração para grupos de instâncias em um cluster em execução. A Amazon EMR só inicia ações de reconfiguração para as classificações que você modifica. Para ter mais informações, consulte [Reconfigurar um grupo de instâncias de um cluster em execução](#).

Classificações do emr-6.5.0

Classificações	Descrição	Ações de reconfiguração
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.	Restarts the ResourceManager service.
container-executor	Altere os valores no arquivo container-executor.cfg YARN do Hadoop.	Not available.
container-log4j	Altere os valores no arquivo container-log4j.properties YARN do Hadoop.	Not available.
core-site	Alterar os valores no arquivo core-site.xml do 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	Alterar as configurações relacionadas ao docker.	Not available.

Classificações	Descrição	Ações de reconfiguração
emrfs-site	Altere EMRFS as configurações.	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	Alterar as configurações do <code>flink-conf.yaml</code> .	Restarts Flink history server.
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.	Restarts Flink history server.
flink-log4j-session	Alterar as configurações de <code>log4j-yarn-session.properties</code> no Flink para sessão do Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Alterar as configurações de <code>log4j-cli.properties</code> no Flink.	Restarts Flink history server.

Classificações	Descrição	Ações de reconfiguração
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do 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	Alterar os valores no arquivo log4j.properties do 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	Alterar a configuração do servidor SSL no Hadoop	Not available.
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop	Not available.

Classificações	Descrição	Ações de reconfiguração
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Mude os valores no HBase ambiente.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.	Not available.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.	This classification should not be reconfigured.
hdfs-env	Mude os valores no HDFS ambiente.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classificações	Descrição	Ações de reconfiguração
hdfs-site	Altere os valores no HDFS hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Mude os valores no HCatalog ambiente.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Configurações EMR selecionadas pela Amazon para o Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.	Not available.

Classificações	Descrição	Ações de reconfiguração
hive-parquet-logging	Alterar os valores no arquivo <code>parquet-logging.properties</code> do Hive.	Not available.
hive-env	Alterar os valores no ambiente do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive <code>schemaTool CLI</code> commands to verify <code>hive-metastore</code> .
hive-exec-log4j2	Altere os valores no arquivo <code>hive-exec-log 4j2.properties</code> do Hive.	Not available.
hive-llap-daemon-log4j2	Altere os valores no arquivo <code>llap-daemon-log 4j2.properties</code> do Hive.	Not available.
hive-log4j2	Alterar os valores no arquivo <code>hive-log4j2.properties</code> do Hive.	Not available.
hive-site	Alterar os valores no arquivo <code>hive-site.xml</code> do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive <code>schemaTool CLI</code> commands to verify <code>hive-metastore</code> . Also restarts Oozie and Zeppelin.
hiveserver2-site	Alterar os valores no arquivo <code>hiveserver2-site.xml</code> do Hive Server2.	Not available.
hue-ini	Alterar os valores no arquivo <code>ini</code> do Hue	Restarts Hue. Also activates Hue <code>config override CLI</code> commands to pick up new configurations.

Classificações	Descrição	Ações de reconfiguração
httpfs-env	Mude os valores no HTTPFS ambiente.	Restarts Hadoop Httpfs service.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.	Not available.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.	Not available.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Altere os valores no ambiente do Hudi.	Not available.
hudi-defaults	Alterar os valores no arquivo hudi-defaults.conf do Hudi.	Not available.
iceberg-defaults	Alterar os valores no arquivo iceberg-defaults.conf do Iceberg.	Not available.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.	Not available.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.	Not available.

Classificações	Descrição	Ações de reconfiguração
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.	Not available.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.	Not available.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.	Restarts Livy Server.
livy-env	Alterar os valores no ambiente do Livy.	Restarts Livy Server.
livy-log4j	Alterar as configurações de log4j.properties no Livy.	Restarts Livy Server.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Alterar os valores no ambiente do Oozie.	Restarts Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.	Restarts Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.	Not available.

Classificações	Descrição	Ações de reconfiguração
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.	Not available.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.	Not available.
pig-env	Alterar os valores no ambiente do Pig.	Not available.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.	Restarts Oozie.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.	Not available.
presto-log	Alterar os valores no arquivo log.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Alterar os valores no arquivo config.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.	Not available.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Altere os valores no arquivo node.properties do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.	Not available.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.	Not available.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.	Not available.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.	Not available.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.	Not available.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.	Not available.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.	Not available.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.	Not available.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.	Not available.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-redis	Alterar os valores no arquivo <code>redis.properties</code> do Presto.	Not available.
presto-connector-redshift	Alterar os valores no arquivo de propriedades <code>redshift</code> do Presto.	Not available.
presto-connector-tpch	Alterar os valores no arquivo <code>tpch.properties</code> do Presto.	Not available.
presto-connector-tpcds	Alterar os valores no arquivo <code>tpcds.properties</code> do Presto.	Not available.
trino-log	Alterar os valores no arquivo <code>log.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-config	Alterar os valores no arquivo <code>config.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Alterar os valores no arquivo <code>password-authenticator.properties</code> do Trino.	Restarts Trino-Server (for Trino)
trino-env	Alterar os valores no arquivo <code>trino-env.sh</code> do Trino.	Restarts Trino-Server (for Trino)
trino-node	Alterar os valores no arquivo <code>node.properties</code> do Trino.	Not available.
trino-connector-blackhole	Alterar os valores no arquivo <code>blackhole.properties</code> do Trino.	Not available.
trino-connector-cassandra	Alterar os valores no arquivo <code>cassandra.properties</code> do Trino.	Not available.
trino-connector-hive	Alterar os valores no arquivo <code>hive.properties</code> do Trino.	Restarts Trino-Server (for Trino)

Classificações	Descrição	Ações de reconfiguração
trino-connector-jmx	Alterar os valores no arquivo <code>jmx.properties</code> do Trino.	Not available.
trino-connector-kafka	Alterar os valores no arquivo <code>kafka.properties</code> do Trino.	Not available.
trino-connector-localfile	Alterar os valores no arquivo <code>localfile.properties</code> do Trino.	Not available.
trino-connector-memory	Alterar os valores no arquivo <code>memory.properties</code> do Trino.	Not available.
trino-connector-mongodb	Alterar os valores no arquivo <code>mongodb.properties</code> do Trino.	Not available.
trino-connector-mysql	Alterar os valores no arquivo <code>mysql.properties</code> do Trino.	Not available.
trino-connector-postgresql	Alterar os valores no arquivo <code>postgresql.properties</code> do Trino.	Not available.
trino-connector-raptor	Alterar os valores no arquivo <code>raptor.properties</code> do Trino.	Not available.
trino-connector-redis	Alterar os valores no arquivo <code>redis.properties</code> do Trino.	Not available.
trino-connector-redshift	Alterar os valores no arquivo <code>redshift.properties</code> do Trino.	Not available.
trino-connector-tpch	Alterar os valores no arquivo <code>tpch.properties</code> do Trino.	Not available.
trino-connector-tpcds	Alterar os valores no arquivo <code>tpcds.properties</code> do Trino.	Not available.
ranger-kms-dbks-site	Altere os valores no arquivo <code>dbks-site.xml</code> do Ranger. KMS	Restarts Ranger KMS Server.

Classificações	Descrição	Ações de reconfiguração
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Altere os valores no arquivo kms-log4j.properties do Ranger. KMS	Not available.
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.	Not available.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Alterar os valores no ambiente do Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.	Not available.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Alterar os valores no ambiente do Sqoop.	Not available.

Classificações	Descrição	Ações de reconfiguração
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.	Not available.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.	Not available.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.	Restart Oozie and HiveServer2.
yarn-env	Mude os valores no YARN ambiente.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.	Restarts Zeppelin.
zeppelin-site	Alterar as definições da configuração em zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.	Restarts Zookeeper server.

Classificações	Descrição	Ações de reconfiguração
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.	Restarts Zookeeper server.

Amazon EMR versão 6.4.0

Versões 6.4.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Trino](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-6.4.0	emr-6.3.1	emr-6.3.0	emr-6.2.1
AWS SDK para Java	1.12.31	1.11.977	1.11.977	1.11.880
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12.10	2.12.10	2.12.10	2.12.10
AmazonCloudWatchAgent	-	-	-	-

	emr-6.4.0	emr-6.3.1	emr-6.3.0	emr-6.2.1
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
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

	emr-6.4.0	emr-6.3.1	emr-6.3.0	emr-6.2.1
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

Notas da versão 6.4.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 6.4.0 da Amazon. As alterações são referentes à versão 6.3.0.

Data da versão inicial: 20 de setembro de 2021

Data da versão atualizada: 21 de março de 2022

Aplicações compatíveis

- AWS SDK for Java versão 1.12.31
- CloudWatch Sink versão 2.2.0
- Conector do DynamoDB versão 4.16.0
- EMRFS versão 2.47.0
- Amazon EMR Goodies versão 3.2.0
- Amazon EMR Kinesis Connector versão 3.5.0
- Amazon EMR Record Server versão 2.1.0
- Amazon EMR Scripts versão 2.5.0
- Flink versão 1.13.1
- Ganglia versão 3.7.2

- AWS Glue Hive Metastore Client versão 3.3.0
- Hadoop versão 3.2.1-amzn-4
- HBase versão 2.4.4-amzn-0
- HBase-operador-tools 1.1.0
- HCatalog versão 3.1.2-amzn-5
- Hive versão 3.1.2-amzn-5
- Hudi versão 0.8.0-amzn-0
- Hue versão 4.9.0
- JDK Versão Java Corretto-8.302.08.1 (compilação 1.8.0_302-b08)
- JupyterHub versão 1.4.1
- Livy versão 0.7.1-incubating
- MXNet versão 1.8.0
- Oozie versão 5.2.1
- Phoenix versão 5.1.2
- Pig versão 0.17.0
- Presto versão 0.254.1-amzn-0
- Trino versão 359
- Apache Ranger KMS (criptografia transparente multi-master) versão 2.0.0
- ranger-plugins 2.0.1-amzn-0
- ranger-s3-plugin 1.2.0
- SageMaker Spark SDK versão 1.4.1
- Scala versão 2.12.10 (VM de servidor aberto de JDK 64 bits, Java 1.8.0_282)
- Spark versão 3.1.2-amzn-0
- spark-rapids 0.4.1
- Sqoop versão 1.4.7
- TensorFlow versão 2.4.1
- Tez versão 0.9.2
- Zeppelin versão 0.9.0
- Zookeeper versão 3.5.7

- Conectores e drivers: DynamoDB Connector 4.16.0

Novos atributos

- [Escalabilidade gerenciada] Otimização de escalabilidade gerenciada de dados do Spark shuffle
 - Para EMR as versões 5.34.0 e posteriores da Amazon e as EMR versões 6.4.0 e posteriores, a escalabilidade gerenciada agora reconhece os dados do Spark shuffle (dados que o Spark redistribui entre partições para realizar operações específicas). Para obter mais informações sobre operações aleatórias, consulte Como [usar a escalabilidade EMR gerenciada na Amazon no Guia de EMR Gerenciamento EMR da Amazon](#) e no Guia de Programação do [Spark](#).
- Em clusters EMR Amazon habilitados para Apache Ranger, você pode usar o Apache SQL Spark para inserir dados ou atualizar as tabelas de metastore do Apache Hive usando, e. INSERT INTO INSERT OVERWRITE ALTER TABLE Ao usar ALTER TABLE com o SparkSQL, um local de partição deve ser o diretório filho de um local de tabela. Atualmente, EMR a Amazon não suporta a inserção de dados em uma partição em que a localização da partição seja diferente da localização da tabela.
- Presto SQL foi [renomeado para Trino](#).
- Hive: a execução de SELECT consultas simples com LIMIT cláusula é acelerada interrompendo a execução da consulta assim que o número de registros mencionados na LIMIT cláusula for obtido. SELECTConsultas simples são consultas que não têm a cláusula GROUP BY/ORDERby ou consultas que não têm um estágio redutor. Por exemplo, SELECT * from <TABLE> WHERE <Condition> LIMIT <Number>.

Controle de simultaneidade do Hudi

- O Hudi agora oferece suporte ao Optimistic Concurrency Control (OCC), que pode ser aproveitado com operações de gravação, como UPSERT e INSERT para permitir alterações de vários gravadores na mesma tabela Hudi. Isso é no nível do arquivoOCC, então quaisquer dois commits (ou escritores) podem gravar na mesma tabela, se suas alterações não entrarem em conflito. Para obter mais informações, consulte o [Controle de simultaneidade do Hudi](#).
- EMROs clusters da Amazon têm o Zookeeper instalado, que pode ser usado como provedor de bloqueio para. OCC Para facilitar o uso desse recurso, os EMR clusters da Amazon têm as seguintes propriedades pré-configuradas:

```
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
```

Para habilitarOCC, você precisa configurar as seguintes propriedades com suas opções de trabalho Hudi ou no nível do cluster usando as configurações da 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)
```

Monitoramento Hudi: CloudWatch integração com a Amazon para reportar Hudi Metrics

- A Amazon EMR oferece suporte à publicação de Hudi Metrics na Amazon CloudWatch. Isso é habilitado com a definição das seguintes configurações necessárias:

```
hoodie.metrics.on=true
hoodie.metrics.reporter.type=CLOUDWATCH
```

- A seguir, são mostradas as configurações opcionais do Hudi que você pode alterar:

Configuração	Descrição	Valor
hoodie.metrics.cloudwatch.report.period.seconds	Frequência (em segundos) na qual reportar métricas para a Amazon CloudWatch	O valor padrão é 60s, o que é adequado para a resolução padrão de um minuto oferecida pela Amazon CloudWatch
hoodie.metrics.cloudwatch.metric.prefix	Prefixo a ser adicionado ao nome de cada métrica	O valor padrão está vazio (sem prefixo)
hoodie.metrics.cloudwatch.namespace	CloudWatch Namespace da Amazon sob o qual as métricas são publicadas	O valor padrão é Hudi

Configuração	Descrição	Valor
<code>hoodie.metrics.cloudwatch.maxDatumsPerSolicitação</code>	Número máximo de dados a serem incluídos em uma solicitação à Amazon CloudWatch	O valor padrão é 20, que é o mesmo que o CloudWatch padrão da Amazon

Suporte e melhorias nas configurações do Amazon EMR Hudi

- Agora, os clientes podem aproveitar o recurso de EMR configurações API e reconfiguração para configurar as configurações do Hudi no nível do cluster. Um novo suporte à configuração baseado em arquivos foi introduzido por meio de `/etc/hudi/conf/hudi-defaults.conf` nos moldes de outras aplicações, como Spark, Hive etc. EMR configura alguns padrões para melhorar a experiência do usuário:
 - `hoodie.datasource.hive_sync.jdbcurl` está configurado para o servidor Hive do cluster URL e não precisa mais ser especificado. Isso é particularmente útil ao executar um trabalho no modo de cluster do Spark, onde anteriormente era necessário especificar o IP EMR principal da Amazon.
 - configurações HBase específicas, que são úteis para usar o HBase índice com o Hudi.
 - Configuração específica do provedor de bloqueio do Zookeeper, conforme discutido em Controle de simultaneidade, o que facilita o uso do Controle de Concorrência Otimista (). OCC
- Alterações adicionais foram introduzidas para reduzir o número de configurações que você precisa passar e inferir automaticamente sempre que possível:
 - A palavra-chave `partitionBy` pode ser usada para especificar a coluna de partição.
 - Ao habilitar o Hive Sync, não é mais obrigatório passar `HIVE_TABLE_OPT_KEY`, `HIVE_PARTITION_FIELDS_OPT_KEY`, `HIVE_PARTITION_EXTRACTOR_CLASS_OPT_KEY`. Esses valores podem ser deduzidos com base no nome da tabela Hudi e no campo de partição.
 - não é obrigatório passar `KEYGENERATOR_CLASS_OPT_KEY`, que pode ser inferido com base em casos mais simples de `SimpleKeyGenerator` e `ComplexKeyGenerator`.

Advertências do Hudi

- O Hudi não permite execução vetorizada no Hive de tabelas Merge on Read (MoR - Mesclar na leitura) e Bootstrap. Por exemplo, `count(*)` apresenta falha com a tabela do Hudi em tempo real quando `hive.vectorized.execution.enabled` está definido como verdadeiro. Como solução alternativa, você pode desabilitar a leitura vetorizada configurando `hive.vectorized.execution.enabled` como `false`.
- O suporte a vários gravadores não é compatível com o atributo de bootstrap do Hudi.
- O Flink Streamer e o Flink SQL são recursos experimentais nesta versão. Esses atributos não são recomendados para uso em implantações de produção.

Alterações, melhorias e problemas resolvidos

Esta é uma versão para corrigir problemas com o Amazon EMR Scaling quando ele falha na escalabilidade ou redução de escala de um cluster com sucesso ou causa falhas no aplicativo.

- Anteriormente, a reinicialização manual do gerenciador de recursos em um cluster multimestre fazia com que os daemons EMR no cluster da Amazon, como o Zookeeper, recarregassem todos os nós anteriormente desativados ou perdidos no arquivo `znode` do Zookeeper. Isso fez com que os limites padrão fossem excedidos em determinadas situações. A Amazon EMR agora remove os registros de nós desativados ou perdidos há mais de uma hora do arquivo do Zookeeper e os limites internos foram aumentados.
- Corrigido um problema em que as solicitações de escalabilidade falhavam em um cluster grande e altamente utilizado quando os daemons no cluster da EMR Amazon estavam executando atividades de verificação de integridade, como coletar o estado do nó e o estado do YARN nó. HDFS Isso estava acontecendo porque os daemons no cluster não conseguiam comunicar os dados do status de saúde de um nó aos componentes internos da Amazon. EMR
- Daemons EMR no cluster aprimorados para rastrear corretamente os estados dos nós quando os endereços IP são reutilizados para melhorar a confiabilidade durante as operações de escalabilidade.
- [SPARK-29683](#). Corrigido um problema em que ocorriam falhas de trabalho durante a redução da escala verticalmente do cluster, pois o Spark presumia que todos os nós disponíveis estavam na lista de negação.
- [YARN-9011](#). Corrigido um problema em que falhas de trabalho ocorriam devido a uma condição de corrida no YARN descomissionamento quando o cluster tentava aumentar ou diminuir a escala.

- Foi corrigido o problema com falhas de etapas ou tarefas durante a escalabilidade do cluster, garantindo que os estados dos nós fossem sempre consistentes entre os EMR daemons no cluster da Amazon e/. YARN HDFS
- Corrigido um problema em que operações de cluster, como redução de escala e envio de etapas, falhavam em EMR clusters da Amazon habilitados com a autenticação Kerberos. Isso ocorreu porque o daemon no EMR cluster da Amazon não renovou o tíquete Kerberos, que é necessário para se comunicar de forma segura com/em execução no nó primário. HDFS YARN
- Configurando um cluster para corrigir problemas de desempenho do Apache YARN Timeline Server versões 1 e 1.5

As versões 1 e 1.5 do Apache YARN Timeline Server podem causar problemas de desempenho com EMR clusters grandes e muito ativos, especialmente `com.yarn.resourcemanager.system-metrics-publisher.enabled=true`, que é a configuração padrão na Amazon. EMR Um YARN Timeline Server v2 de código aberto resolve o problema de desempenho relacionado à escalabilidade do YARN Timeline Server.

Outras soluções alternativas para esse problema incluem:

- Configurando `yarn.resourcemanager.system-metrics-publisher.enabled=false` em `yarn-site.xml`.
- Habilitar a correção para esse problema na criação de um cluster, conforme descrito abaixo.

As seguintes EMR versões da Amazon contêm uma correção para esse problema de desempenho YARN do Timeline Server.

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

Para habilitar a correção em qualquer uma das EMR versões da Amazon especificadas acima, defina essas propriedades `true` em um JSON arquivo de configurações que é passado usando o [parâmetro de aws emr create-cluster comando](#): `--configurations file:///./configurations.json`. Ou habilite a correção usando a [interface do usuário do console de reconfiguração](#).

Exemplo de conteúdo do arquivo `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": []  
}  
]
```

- Os servidores Web HDFS e HTTPFs estão desativados por padrão. Você pode reativar a Web HDFS usando a configuração do Hadoop, `dfs.webhdfs.enabled`. O servidor HttpFS pode ser iniciado usando `sudo systemctl start hadoop-httpfs`.
- HTTPS agora está habilitado por padrão para os repositórios Amazon Linux. Se você estiver usando uma VPCE política do Amazon S3 para restringir o acesso a buckets específicos, deverá adicionar o novo bucket Amazon Linux ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` à sua política (`$region` substitua pela região onde está o endpoint). Para obter mais informações, consulte esse tópico nos fóruns de AWS discussão. [Anúncio: o Amazon Linux 2 agora oferece suporte à capacidade de uso HTTPS ao se conectar a repositórios de pacotes.](#)
- Hive: o desempenho da consulta de gravação foi aprimorado ao permitir o uso de um diretório temporário ativado HDFS para o último trabalho. Os dados temporários do trabalho final são gravados no Amazon S3, e o desempenho é aprimorado porque os dados são movidos HDFS para o local final da tabela (Amazon S3) em vez de entre dispositivos Amazon S3. HDFS
- Hive: melhoria do tempo de compilação de consultas em até 2,5 vezes com a remoção de partições de metastores do Glue.
- Por padrão, quando UDFs os integrados são passados pelo Hive para o Hive Metastore Server, somente um subconjunto desses incorporados é UDFs passado para o Glue Metastore, já que o Glue suporta apenas operadores de expressão limitados. Se você definir `hive.glue.partition.pruning.client=true`, toda a remoção de partições ocorrerá no lado do cliente. Se você definir `hive.glue.partition.pruning.server=true`, toda a remoção de partições ocorrerá no lado do servidor.

Problemas conhecidos

- As consultas Hue não funcionam no Amazon EMR 6.4.0 porque o servidor Apache Hadoop HTTPFs está desativado por padrão. Para usar o Hue na Amazon EMR 6.4.0, inicie manualmente o servidor HTTPFs no nó EMR primário da Amazon usando `sudo systemctl start hadoop-httpfs` ou use [uma etapa da Amazon](#). EMR
- O recurso Amazon EMR Notebooks usado com a representação de usuário do Livy não funciona porque o HTTPFs está desativado por padrão. Nesse caso, o EMR notebook não pode se conectar ao cluster que tem a representação Livy ativada. A solução alternativa é iniciar o servidor HTTPFs

antes de conectar o EMR notebook ao cluster usando o. `sudo systemctl start hadoop-httpfs`

- Na EMR versão 6.4.0 da Amazon, o Phoenix não oferece suporte ao componente de conectores Phoenix.
- Para usar as ações do Spark com o Apache Oozie, você deve adicionar a seguinte configuração ao seu arquivo `workflow.xml` do Oozie. Caso contrário, várias bibliotecas críticas, como a Hadoop e a Hadoop, EMRFS estarão ausentes do classpath dos executores do Spark que o Oozie lança.

```
<spark-opts>--conf spark.yarn.populateHadoopClasspath=true</spark-opts>
```

- Quando você usa o Spark com a formatação de localização de partições do Hive para ler dados no Amazon S3 e executa o Spark nas EMR versões 5.30.0 a 5.36.0 e 6.2.0 a 6.9.0 da Amazon, você pode encontrar um problema que impede seu cluster de ler os dados corretamente. Isso poderá acontecer se suas partições tiverem todas as características a seguir:
 - Duas ou mais partições são verificadas na mesma tabela.
 - Pelo menos um caminho de diretório de partição é um prefixo de pelo menos outro caminho de diretório de partição, por exemplo, `s3://bucket/table/p=a` é um prefixo de `s3://bucket/table/p=a b`.
 - O primeiro caractere que segue o prefixo no outro diretório de partição tem um valor UTF -8 menor que o / caractere (U+002F). Por exemplo, o caractere de espaço (U+0020) que ocorre entre a e b em `s3://bucket/table/p=a b` se enquadra nessa categoria. Observe que existem 14 outros caracteres que não são de controle: `!"#$%&'()*+,-.` Para obter mais informações, consulte a [tabela de codificação UTF -8 e caracteres Unicode](#).

Como solução alternativa para esse problema, defina a configuração `spark.sql.sources.fastS3PartitionDiscovery.enabled` como `false` na classificação `spark-defaults`.

Versões de componente 6.4.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são

a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado myapp-component com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como 2.2-amzn-2.

Componente	Version (Versão)	Descrição
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK
emr-ddb	4.16.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	3.2.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.5.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-notebook-env	1.3.0	Ambiente Conda para bloco de anotações do emr, que inclui o jupyter enterprise gateway
emr-s3-dist-cp	2.18.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	2.1.0	EMRConector S3Select
emrfs	2.47.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
flink-client	1.13.1	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
flink-jobmanager-config	1.13.1	Gerenciando recursos em EMR nós para o Apache JobManager Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	3.2.1-amzn-4	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	3.2.1-amzn-4	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	3.2.1-amzn-4	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	3.2.1-amzn-4	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.

Componente	Version (Versão)	Descrição
hadoop-hdfs-journalnode	3.2.1-amzn-4	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httfs-server	3.2.1-amzn-4	HTTPendpoint para HDFS operações.
hadoop-kms-server	3.2.1-amzn-4	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	3.2.1-amzn-4	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	3.2.1-amzn-4	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	3.2.1-amzn-4	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	3.2.1-amzn-4	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	2.4.4-amzn-0	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	2.4.4-amzn-0	Serviço para atender a uma ou mais HBase regiões.

Componente	Version (Versão)	Descrição
hbase-client	2.4.4-amzn-0	HBase cliente de linha de comando.
hbase-rest-server	2.4.4-amzn-0	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	2.4.4-amzn-0	Serviço que fornece um endpoint Thrift para HBase
hcatalog-client	3.1.2-amzn-5	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	3.1.2-amzn-5	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	3.1.2-amzn-5	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	3.1.2-amzn-5	O cliente da linha de comando do Hive.
hive-hbase	3.1.2-amzn-5	Cliente do Hive-hbase.
hive-metastore-server	3.1.2-amzn-5	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL

Componente	Version (Versão)	Descrição
hive-server2	3.1.2-amzn-5	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.8.0-amzn-0	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.
hudi-presto	0.8.0-amzn-0	Biblioteca de pacotes para executar o Presto com o Hudi.
hudi-trino	0.8.0-amzn-0	Biblioteca de pacotes para execução do Presto com o Hudi.
hudi-spark	0.8.0-amzn-0	Biblioteca de pacotes para executar o Spark com o Hudi.
hue-server	4.9.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	1.4.1	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.7.1-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mxnet	1.8.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.

Componente	Version (Versão)	Descrição
mariadb-server	5.5.68+	Servidor de banco de dados MariaDB.
nvidia-cuda	10.1.243	Drivers NVIDIA e toolkit CUDA
oozie-client	5.2.1	O cliente da linha de comando do Oozie.
oozie-server	5.2.1	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	4.5.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	5.1.2	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	5.1.2	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.254.1-amzn-0	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.254.1-amzn-0	O serviço que executa partes de uma consulta.
presto-client	0.254.1-amzn-0	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.

Componente	Version (Versão)	Descrição
trino-coordinator	359	Serviço que aceita consultas e gerencia a execução de consultas entre trino-workers.
trino-worker	359	O serviço que executa partes de uma consulta.
trino-client	359	Cliente de linha de comando do Trino instalado em mestres de espera de um cluster de alta disponibilidade (HA) em que o servidor do Trino não é iniciado.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	4.0.2	O projeto R para computação estatística
ranger-kms-server	2.0.0	Sistema de gerenciamento de chaves Apache Ranger
spark-client	3.1.2-amzn-0	Os clientes da linha de comando do Spark.
spark-history-server	3.1.2-amzn-0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	3.1.2-amzn-0	Mecanismo de execução na memória para YARN.

Componente	Version (Versão)	Descrição
spark-yarn-slave	3.1.2-amzn-0	Bibliotecas do Apache Spark necessárias para escravos. YARN
spark-rapids	0.4.1	RAPIDSPlugin Nvidia Spark que acelera o Apache Spark com o. GPUs
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	2.4.1	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.2	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.41+	HTTPServidor Apache.
zeppelin-server	0.9.0	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.5.7	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.5.7	ZooKeeper cliente de linha de comando.

Classificações de configuração 6.4.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Ocorrem ações de reconfiguração quando você especifica uma configuração para grupos de instâncias em um cluster em execução. A Amazon EMR só inicia ações de reconfiguração para as classificações que você modifica. Para ter mais informações, consulte [Reconfigurar um grupo de instâncias de um cluster em execução](#).

Classificações do emr-6.4.0

Classificações	Descrição	Ações de reconfiguração
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.	Restarts the Resource Manager service.
container-executor	Altere os valores no arquivo <code>container-executor.cfg</code> YARN do Hadoop.	Not available.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.	Not available.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do 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

Classificações	Descrição	Ações de reconfiguração
		Httpfs, and MapReduce-HistoryServer.
docker-conf	Alterar as configurações relacionadas ao docker.	Not available.
emrfs-site	Altere EMRFS as configurações.	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	Alterar as configurações do flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Alterar as configurações de log4j.properties no Flink.	Restarts Flink history server.
flink-log4j-session	Alterar as configurações de log4j-yarn-session.properties no Flink para sessão do Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.	Restarts Flink history server.

Classificações	Descrição	Ações de reconfiguração
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do 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	Alterar os valores no arquivo log4j.properties do 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	Alterar a configuração do servidor SSL no Hadoop	Not available.
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop	Not available.

Classificações	Descrição	Ações de reconfiguração
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Mude os valores no HBase ambiente.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.	Not available.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.	This classification should not be reconfigured.
hdfs-env	Mude os valores no HDFS ambiente.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classificações	Descrição	Ações de reconfiguração
hdfs-site	Altere os valores no HDFS hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Mude os valores no HCatalog ambiente.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Configurações EMR selecionadas pela Amazon para o Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.	Not available.

Classificações	Descrição	Ações de reconfiguração
hive-parquet-logging	Alterar os valores no arquivo <code>parquet-logging.properties</code> do Hive.	Not available.
hive-env	Alterar os valores no ambiente do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive <code>schemaTool CLI</code> commands to verify <code>hive-metastore</code> .
hive-exec-log4j2	Altere os valores no arquivo <code>hive-exec-log 4j2.properties</code> do Hive.	Not available.
hive-llap-daemon-log4j2	Altere os valores no arquivo <code>llap-daemon-log 4j2.properties</code> do Hive.	Not available.
hive-log4j2	Alterar os valores no arquivo <code>hive-log4j2.properties</code> do Hive.	Not available.
hive-site	Alterar os valores no arquivo <code>hive-site.xml</code> do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive <code>schemaTool CLI</code> commands to verify <code>hive-metastore</code> . Also restarts Oozie and Zeppelin.
hiveserver2-site	Alterar os valores no arquivo <code>hiveserver2-site.xml</code> do Hive Server2.	Not available.
hue-ini	Alterar os valores no arquivo <code>ini</code> do Hue	Restarts Hue. Also activates Hue <code>config override CLI</code> commands to pick up new configurations.

Classificações	Descrição	Ações de reconfiguração
httpfs-env	Mude os valores no HTTPFS ambiente.	Restarts Hadoop Httpfs service.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.	Not available.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.	Not available.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Altere os valores no ambiente do Hudi.	Not available.
hudi-defaults	Alterar os valores no arquivo hudi-defaults.conf do Hudi.	Not available.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.	Not available.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.	Not available.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.	Not available.

Classificações	Descrição	Ações de reconfiguração
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.	Not available.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.	Restarts Livy Server.
livy-env	Alterar os valores no ambiente do Livy.	Restarts Livy Server.
livy-log4j	Alterar as configurações de log4j.properties no Livy.	Restarts Livy Server.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Alterar os valores no ambiente do Oozie.	Restarts Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.	Restarts Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.	Not available.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.	Not available.

Classificações	Descrição	Ações de reconfiguração
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.	Not available.
pig-env	Alterar os valores no ambiente do Pig.	Not available.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.	Restarts Oozie.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.	Not available.
presto-log	Alterar os valores no arquivo log.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Alterar os valores no arquivo config.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.	Not available.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Altere os valores no arquivo node.properties do Presto.	Not available.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-cassandra	Alterar os valores no arquivo <code>cassandra.properties</code> do Presto.	Not available.
presto-connector-hive	Alterar os valores no arquivo <code>hive.properties</code> do Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Alterar os valores no arquivo <code>jmx.properties</code> do Presto.	Not available.
presto-connector-kafka	Alterar os valores no arquivo <code>kafka.properties</code> do Presto.	Not available.
presto-connector-localfile	Alterar os valores no arquivo <code>localfile.properties</code> do Presto.	Not available.
presto-connector-memory	Alterar os valores no arquivo <code>memory.properties</code> do Presto.	Not available.
presto-connector-mongodb	Alterar os valores no arquivo <code>mongodb.properties</code> do Presto.	Not available.
presto-connector-mysql	Alterar os valores no arquivo <code>mysql.properties</code> do Presto.	Not available.
presto-connector-postgresql	Alterar os valores no arquivo <code>postgresql.properties</code> do Presto.	Not available.
presto-connector-raptor	Alterar os valores no arquivo <code>raptor.properties</code> do Presto.	Not available.
presto-connector-redis	Alterar os valores no arquivo <code>redis.properties</code> do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.	Not available.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.	Not available.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.	Not available.
trino-log	Alterar os valores no arquivo log.properties do Trino.	Restarts Trino-Server (for Trino)
trino-config	Alterar os valores no arquivo config.properties do Trino.	Restarts Trino-Server (for Trino)
trino-password-authenticator	Alterar os valores no arquivo password-authenticator.properties do Trino.	Restarts Trino-Server (for Trino)
trino-env	Alterar os valores no arquivo trino-env.sh do Trino.	Restarts Trino-Server (for Trino)
trino-node	Alterar os valores no arquivo node.properties do Trino.	Not available.
trino-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Trino.	Not available.
trino-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Trino.	Not available.
trino-connector-hive	Alterar os valores no arquivo hive.properties do Trino.	Restarts Trino-Server (for Trino)
trino-connector-jmx	Alterar os valores no arquivo jmx.properties do Trino.	Not available.

Classificações	Descrição	Ações de reconfiguração
trino-connector-kafka	Alterar os valores no arquivo kafka.properties do Trino.	Not available.
trino-connector-localfile	Alterar os valores no arquivo localfile.properties do Trino.	Not available.
trino-connector-memory	Alterar os valores no arquivo memory.properties do Trino.	Not available.
trino-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Trino.	Not available.
trino-connector-mysql	Alterar os valores no arquivo mysql.properties do Trino.	Not available.
trino-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Trino.	Not available.
trino-connector-raptor	Alterar os valores no arquivo raptor.properties do Trino.	Not available.
trino-connector-redis	Alterar os valores no arquivo redis.properties do Trino.	Not available.
trino-connector-redshift	Alterar os valores no arquivo redshift.properties do Trino.	Not available.
trino-connector-tpch	Alterar os valores no arquivo tpch.properties do Trino.	Not available.
trino-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Trino.	Not available.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS	Restarts Ranger KMS Server.

Classificações	Descrição	Ações de reconfiguração
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Altere os valores no arquivo kms-log4j.properties do Ranger. KMS	Not available.
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.	Not available.
spark	Configurações EMR seleciona das pela Amazon para o Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Alterar os valores no ambiente do Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.	Not available.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Alterar os valores no ambiente do Sqoop.	Not available.

Classificações	Descrição	Ações de reconfiguração
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.	Not available.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.	Not available.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.	Restart Oozie and HiveServer2.
yarn-env	Mude os valores no YARN ambiente.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.	Restarts Zeppelin.
zeppelin-site	Alterar as definições da configuração em zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.	Restarts Zookeeper server.

Classificações	Descrição	Ações de reconfiguração
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.	Restarts Zookeeper server.

Amazon EMR versão 6.3.1

Versões 6.3.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [PrestoSQL](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-6.3.1	emr-6.3.0	emr-6.2.1	emr-6.2.0
AWS SDK para Java	1.11.977	1.11.977	1.11.880	1.11.880
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12.10	2.12.10	2.12.10	2.12.10
AmazonCloudWatchAgent	-	-	-	-

	emr-6.3.1	emr-6.3.0	emr-6.2.1	emr-6.2.0
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
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

	emr-6.3.1	emr-6.3.0	emr-6.2.1	emr-6.2.0
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

Notas da versão 6.3.1

Esta é uma versão para corrigir problemas com o Amazon EMR Scaling quando ele falha na escalabilidade ou redução de escala de um cluster com sucesso ou causa falhas no aplicativo.

Alterações, melhorias e problemas resolvidos

- Corrigido um problema em que as solicitações de escalabilidade falhavam em um cluster grande e altamente utilizado quando os daemons no cluster da EMR Amazon estavam executando atividades de verificação de integridade, como coletar o estado do nó e o estado do YARN nó. HDFS Isso estava acontecendo porque os daemons no cluster não conseguiam comunicar os dados do status de saúde de um nó aos componentes internos da Amazon. EMR
- Daemons EMR no cluster aprimorados para rastrear corretamente os estados dos nós quando os endereços IP são reutilizados para melhorar a confiabilidade durante as operações de escalabilidade.
- [SPARK-29683](#). Corrigido um problema em que ocorriam falhas de trabalho durante a redução da escala verticalmente do cluster, pois o Spark presumia que todos os nós disponíveis estavam na lista de negação.
- [YARN-9011](#). Corrigido um problema em que falhas de trabalho ocorriam devido a uma condição de corrida no YARN descomissionamento quando o cluster tentava aumentar ou diminuir a escala.
- Foi corrigido o problema com falhas de etapas ou tarefas durante a escalabilidade do cluster, garantindo que os estados dos nós fossem sempre consistentes entre os EMR daemons no cluster da Amazon e/. YARN HDFS

- Corrigido um problema em que operações de cluster, como redução de escala e envio de etapas, falhavam em EMR clusters da Amazon habilitados com a autenticação Kerberos. Isso ocorreu porque o daemon no EMR cluster da Amazon não renovou o tíquete Kerberos, que é necessário para se comunicar de forma segura com/em execução no nó primário. HDFS YARN
- HTTPS agora está habilitado por padrão para os repositórios Amazon Linux. Se você estiver usando uma VPCE política do Amazon S3 para restringir o acesso a buckets específicos, deverá adicionar o novo bucket Amazon Linux ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` à sua política (`$region` substitua pela região onde está o endpoint). Para obter mais informações, consulte esse tópico nos fóruns de AWS discussão. [Anúncio: o Amazon Linux 2 agora oferece suporte à capacidade de uso HTTPS ao se conectar a repositórios de pacotes.](#)

Problemas conhecidos

- Quando você usa o Spark com a formatação de localização de partições do Hive para ler dados no Amazon S3 e executa o Spark nas EMR versões 5.30.0 a 5.36.0 e 6.2.0 a 6.9.0 da Amazon, você pode encontrar um problema que impede seu cluster de ler os dados corretamente. Isso poderá acontecer se suas partições tiverem todas as características a seguir:
 - Duas ou mais partições são verificadas na mesma tabela.
 - Pelo menos um caminho de diretório de partição é um prefixo de pelo menos outro caminho de diretório de partição, por exemplo, `s3://bucket/table/p=a` é um prefixo de `s3://bucket/table/p=a b`.
 - O primeiro caractere que segue o prefixo no outro diretório de partição tem um valor UTF-8 menor que o / caractere (U+002F). Por exemplo, o caractere de espaço (U+0020) que ocorre entre a e b em `s3://bucket/table/p=a b` se enquadra nessa categoria. Observe que existem 14 outros caracteres que não são de controle: `!"#$%&'()*+,-.` Para obter mais informações, consulte a [tabela de codificação UTF-8 e caracteres Unicode](#).

Como solução alternativa para esse problema, defina a configuração `spark.sql.sources.fastS3PartitionDiscovery.enabled` como `false` na classificação `spark-defaults`.

Versões de componente 6.3.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou

aws. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado myapp-component com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como 2.2-amzn-2.

Componente	Version (Versão)	Descrição
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK
emr-ddb	4.16.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	3.2.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.5.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-notebook-env	1.2.0	Ambiente Conda para bloco de anotações do emr, que inclui o jupyter enterprise gateway
emr-s3-dist-cp	2.18.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	2.1.0	EMRConector S3Select
emrfs	2.46.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
flink-client	1.12.1	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
flink-jobmanager-config	1.12.1	Gerenciando recursos em EMR nós para o Apache JobManager Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	3.2.1-amzn-3.1	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	3.2.1-amzn-3.1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	3.2.1-amzn-3.1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	3.2.1-amzn-3.1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.

Componente	Version (Versão)	Descrição
hadoop-hdfs-journalnode	3.2.1-amzn-3.1	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httfs-server	3.2.1-amzn-3.1	HTTPendpoint para HDFS operações.
hadoop-kms-server	3.2.1-amzn-3.1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	3.2.1-amzn-3.1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	3.2.1-amzn-3.1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	3.2.1-amzn-3.1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	3.2.1-amzn-3.1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	2.2.6-amzn-1	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	2.2.6-amzn-1	Serviço para atender a uma ou mais HBase regiões.

Componente	Version (Versão)	Descrição
hbase-client	2.2.6-amzn-1	HBase cliente de linha de comando.
hbase-rest-server	2.2.6-amzn-1	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	2.2.6-amzn-1	Serviço que fornece um endpoint Thrift para HBase.
hcatalog-client	3.1.2-amzn-4	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	3.1.2-amzn-4	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	3.1.2-amzn-4	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	3.1.2-amzn-4	O cliente da linha de comando do Hive.
hive-hbase	3.1.2-amzn-4	Cliente do Hive-hbase.
hive-metastore-server	3.1.2-amzn-4	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL

Componente	Version (Versão)	Descrição
hive-server2	3.1.2-amzn-4	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.7.0-amzn-0	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.
hudi-presto	0.7.0-amzn-0	Biblioteca de pacotes para executar o Presto com o Hudi.
hudi-prestosql	0.7.0-amzn-0	Biblioteca de pacotes para executar o Presto SQL com o Hudi.
hudi-spark	0.7.0-amzn-0	Biblioteca de pacotes para executar o Spark com o Hudi.
hue-server	4.9.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	1.2.2	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.7.0-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mxnet	1.7.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.

Componente	Version (Versão)	Descrição
mariadb-server	5.5.68+	Servidor de banco de dados MariaDB.
nvidia-cuda	10.1.243	Drivers NVIDIA e toolkit CUDA
oozie-client	5.2.1	O cliente da linha de comando do Oozie.
oozie-server	5.2.1	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	4.5.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	5.0.0- HBase -2,0	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	5.0.0- HBase -2,0	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.245.1-amzn-0	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.245.1-amzn-0	O serviço que executa partes de uma consulta.
presto-client	0.245.1-amzn-0	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.

Componente	Version (Versão)	Descrição
prestosql-coordinator	350	O serviço que aceita consultas e gerencia a execução de consultas entre os prestosql-workers.
prestosql-worker	350	O serviço que executa partes de uma consulta.
prestosql-client	350	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	4.0.2	O projeto R para computação estatística
ranger-kms-server	2.0.0	Sistema de gerenciamento de chaves Apache Ranger
spark-client	3.1.1-amzn-0.1	Os clientes da linha de comando do Spark.
spark-history-server	3.1.1-amzn-0.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	3.1.1-amzn-0.1	Mecanismo de execução na memória para YARN.

Componente	Version (Versão)	Descrição
spark-yarn-slave	3.1.1-amzn-0.1	Bibliotecas do Apache Spark necessárias para escravos. YARN
spark-rapids	0.4.1	RAPIDSPlugin Nvidia Spark que acelera o Apache Spark com o. GPUs
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	2.4.1	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.2	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.41+	HTTPServidor Apache.
zeppelin-server	0.9.0	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.14	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.14	ZooKeeper cliente de linha de comando.

Classificações de configuração 6.3.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Ocorrem ações de reconfiguração quando você especifica uma configuração para grupos de instâncias em um cluster em execução. A Amazon EMR só inicia ações de reconfiguração para as classificações que você modifica. Para ter mais informações, consulte [Reconfigurar um grupo de instâncias de um cluster em execução](#).

Classificações do emr-6.3.1

Classificações	Descrição	Ações de reconfiguração
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.	Restarts the Resource Manager service.
container-executor	Altere os valores no arquivo <code>container-executor.cfg</code> YARN do Hadoop.	Not available.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.	Not available.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do 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

Classificações	Descrição	Ações de reconfiguração
		Httpfs, and MapReduce-HistoryServer.
docker-conf	Alterar as configurações relacionadas ao docker.	Not available.
emrfs-site	Altere EMRFS as configurações.	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	Alterar as configurações do flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Alterar as configurações de log4j.properties no Flink.	Restarts Flink history server.
flink-log4j-session	Alterar as configurações de log4j-yarn-session.properties no Flink para sessão do Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.	Restarts Flink history server.

Classificações	Descrição	Ações de reconfiguração
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do 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	Alterar os valores no arquivo log4j.properties do 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	Alterar a configuração do servidor SSL no Hadoop	Not available.
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop	Not available.

Classificações	Descrição	Ações de reconfiguração
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Mude os valores no HBase ambiente.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.	Not available.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.	This classification should not be reconfigured.
hdfs-env	Mude os valores no HDFS ambiente.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classificações	Descrição	Ações de reconfiguração
hdfs-site	Altere os valores no HDFS hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Mude os valores no HCatalog ambiente.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Configurações EMR selecionadas pela Amazon para o Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.	Not available.

Classificações	Descrição	Ações de reconfiguração
hive-parquet-logging	Alterar os valores no arquivo <code>parquet-logging.properties</code> do Hive.	Not available.
hive-env	Alterar os valores no ambiente do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive <code>schemaTool CLI</code> commands to verify <code>hive-metastore</code> .
hive-exec-log4j2	Altere os valores no arquivo <code>hive-exec-log 4j2.properties</code> do Hive.	Not available.
hive-llap-daemon-log4j2	Altere os valores no arquivo <code>llap-daemon-log 4j2.properties</code> do Hive.	Not available.
hive-log4j2	Alterar os valores no arquivo <code>hive-log4j2.properties</code> do Hive.	Not available.
hive-site	Alterar os valores no arquivo <code>hive-site.xml</code> do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive <code>schemaTool CLI</code> commands to verify <code>hive-metastore</code> . Also restarts Oozie and Zeppelin.
hiveserver2-site	Alterar os valores no arquivo <code>hiveserver2-site.xml</code> do Hive Server2.	Not available.
hue-ini	Alterar os valores no arquivo <code>ini</code> do Hue	Restarts Hue. Also activates Hue <code>config override CLI</code> commands to pick up new configurations.

Classificações	Descrição	Ações de reconfiguração
httpfs-env	Mude os valores no HTTPFS ambiente.	Restarts Hadoop Httpfs service.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.	Not available.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.	Not available.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Altere os valores no ambiente do Hudi.	Not available.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.	Not available.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.	Not available.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.	Not available.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.	Not available.

Classificações	Descrição	Ações de reconfiguração
livy-conf	Alterar os valores no arquivo livy.conf do Livy.	Restarts Livy Server.
livy-env	Alterar os valores no ambiente do Livy.	Restarts Livy Server.
livy-log4j	Alterar as configurações de log4j.properties no Livy.	Restarts Livy Server.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Alterar os valores no ambiente do Oozie.	Restarts Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.	Restarts Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.	Not available.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.	Not available.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.	Restarts Phoenix-QueryServer.

Classificações	Descrição	Ações de reconfiguração
phoenix-metrics	Alterar os valores no arquivo <code>hadoop-metrics2-phoenix.properties</code> do Phoenix.	Not available.
pig-env	Alterar os valores no ambiente do Pig.	Not available.
pig-properties	Alterar os valores no arquivo <code>pig.properties</code> do Pig.	Restarts Oozie.
pig-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Pig.	Not available.
presto-log	Alterar os valores no arquivo <code>log.properties</code> do Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Alterar os valores no arquivo <code>config.properties</code> do Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Altere os valores no arquivo <code>password-authenticator.properties</code> do Presto.	Not available.
presto-env	Altere os valores no arquivo <code>presto-env.sh</code> do Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Altere os valores no arquivo <code>node.properties</code> do Presto.	Not available.
presto-connector-blackhole	Alterar os valores no arquivo <code>blackhole.properties</code> do Presto.	Not available.
presto-connector-cassandra	Alterar os valores no arquivo <code>cassandra.properties</code> do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.	Not available.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.	Not available.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.	Not available.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.	Not available.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.	Not available.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.	Not available.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.	Not available.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.	Not available.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.	Not available.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.	Not available.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.	Not available.
prestosql-log	Alterar os valores no arquivo log.properties do Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-config	Alterar os valores no arquivo config.properties do Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-env	Altere os valores no arquivo presto-env.sh do Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-node	Altere os valores no arquivo SQL node.properties do Presto.	Not available.
prestosql-connector-blackhole	Altere os valores no arquivo SQL blackhole.properties do Presto.	Not available.
prestosql-connector-cassandra	Altere os valores no arquivo SQL cassandra.properties do Presto.	Not available.
prestosql-connector-hive	Altere os valores no arquivo SQL hive.properties do Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-connector-jmx	Altere os valores no arquivo SQL jmx.properties do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
prestosql-connector-kafka	Altere os valores no arquivo SQL kafka.properties do Presto.	Not available.
prestosql-connector-localfile	Altere os valores no arquivo SQL localfile.properties do Presto.	Not available.
prestosql-connector-memory	Altere os valores no arquivo SQL memory.properties do Presto.	Not available.
prestosql-connector-mongodb	Altere os valores no arquivo SQL mongodb.properties do Presto.	Not available.
prestosql-connector-mysql	Altere os valores no arquivo SQL mysql.properties do Presto.	Not available.
prestosql-connector-postgresql	Altere os valores no arquivo SQL postgresql.properties do Presto.	Not available.
prestosql-connector-raptor	Altere os valores no arquivo SQL raptor.properties do Presto.	Not available.
prestosql-connector-redis	Altere os valores no arquivo SQL redis.properties do Presto.	Not available.
prestosql-connector-redshift	Altere os valores no arquivo SQL redshift.properties do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
prestosql-connector-tpch	Altere os valores no arquivo SQL tpch.properties do Presto.	Not available.
prestosql-connector-tpcds	Altere os valores no arquivo SQL tpcds.properties do Presto.	Not available.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Altere os valores no arquivo kms-log4j.properties do Ranger. KMS	Not available.
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.	Not available.
spark	Configurações EMR seleciona das pela Amazon para o Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Alterar os valores no ambiente do Spark.	Restarts Spark history server and Spark thrift server.

Classificações	Descrição	Ações de reconfiguração
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.	Not available.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Alterar os valores no ambiente do Sqoop.	Not available.
sqoop-oracoop-site	Altere os valores no arquivo oracoop-site.xml OraOop do Sqoop.	Not available.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.	Not available.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.	Restart Oozie and HiveServer2.
yarn-env	Mude os valores no YARN ambiente.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.

Classificações	Descrição	Ações de reconfiguração
zeppelin-env	Alterar os valores no ambiente do Zeppelin.	Restarts Zeppelin.
zeppelin-site	Alterar as definições da configuração em zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.	Restarts Zookeeper server.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.	Restarts Zookeeper server.

Amazon EMR versão 6.3.0

Versões 6.3.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [PrestoSQL](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-6.3.0	emr-6.2.1	emr-6.2.0	emr-6.1.1
AWS SDK para Java	1.11.977	1.11.880	1.11.880	1.11.828
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12.10	2.12.10	2.12.10	2.12.10
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

Notas da versão 6.3.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 6.3.0 da Amazon. As alterações são referentes à versão 6.2.0.

Data da versão inicial: 12 de maio de 2021

Data da última atualização: 9 de agosto de 2021

Aplicações compatíveis

- AWS SDK for Java versão 1.11.977

- CloudWatch Sink versão 2.1.0
- Conector do DynamoDB versão 4.16.0
- EMRFS versão 2.46.0
- Amazon EMR Goodies versão 3.2.0
- Amazon EMR Kinesis Connector versão 3.5.0
- Amazon EMR Record Server versão 2.0.0
- Amazon EMR Scripts versão 2.5.0
- Flink versão 1.12.1
- Ganglia versão 3.7.2
- AWS Glue Hive Metastore Client versão 3.2.0
- Hadoop versão 3.2.1-amzn-3
- HBase versão 2.2.6-amzn-1
- HBase-operator-tools 1.0.0
- HCatalog versão 3.1.2-amzn-0
- Hive versão 3.1.2-amzn-4
- Hudi versão 0.7.0-amzn-0
- Hue versão 4.9.0
- JDK Versão Java Corretto-8.282.08.1 (compilação 1.8.0_282-b08)
- JupyterHub versão 1.2.0
- Livy versão 0.7.0-incubating
- MXNet versão 1.7.0
- Oozie versão 5.2.1
- Phoenix versão 5.0.0
- Pig versão 0.17.0
- Presto versão 0.245.1-amzn-0
- Presto SQL versão 350
- Apache Ranger KMS (criptografia transparente multi-master) versão 2.0.0
- ranger-plugins 2.0.1-amzn-0

- ranger-s3-plugin 1.1.0
- SageMaker Spark SDK versão 1.4.1
- Scala versão 2.12.10 (VM de servidor aberto de JDK 64 bits, Java 1.8.0_282)
- Spark versão 3.1.1-amzn-0
- spark-rapids 0.4.1
- Sqoop versão 1.4.7
- TensorFlow versão 2.4.1
- Tez versão 0.9.2
- Zeppelin versão 0.9.0
- Zookeeper versão 3.4.14
- Conectores e drivers: DynamoDB Connector 4.16.0

Novos atributos

- A Amazon EMR oferece suporte aos pontos de acesso do Amazon S3, um recurso do Amazon S3 que permite gerenciar facilmente o acesso a data lakes compartilhados. Usando o alias do Amazon S3 Access Point, você pode simplificar o acesso aos dados em grande escala na Amazon. EMR Você pode usar os pontos de acesso do Amazon S3 com todas as versões da Amazon sem EMR custo adicional em todas as AWS regiões onde a Amazon EMR está disponível. Para saber mais sobre Pontos de Acesso Amazon S3 e aliases de ponto de acesso, consulte [Usar um alias em estilo de bucket para seu ponto de acesso](#) no Guia do usuário do Amazon S3.
- Novos `ListReleaseLabel` API parâmetros `DescribeReleaseLabel` e parâmetros fornecem detalhes da etiqueta EMR de lançamento da Amazon. Você pode listar programaticamente os lançamentos disponíveis na região em que a API solicitação é executada e listar os aplicativos disponíveis para uma etiqueta de EMR lançamento específica da Amazon. Os parâmetros da etiqueta de lançamento também listam as EMR versões da Amazon que oferecem suporte a um aplicativo específico, como o Spark. Essas informações podem ser usadas para iniciar programaticamente clusters da AmazonEMR. Por exemplo, você pode iniciar um cluster usando a versão mais recente dos resultados de `ListReleaseLabel`. Para obter mais informações, consulte [DescribeReleaseLabel](#) e [ListReleaseLabels](#) na EMRAPIReferência da Amazon.
- Com o Amazon EMR 6.3.0, você pode iniciar um cluster que se integra nativamente com o Apache Ranger. O Apache Ranger é uma estrutura de código aberto para habilitar, monitorar e gerenciar

uma segurança de dados abrangente em toda a plataforma Hadoop. Para obter mais informações, consulte [Apache Ranger](#). Com a integração nativa, você pode trazer seu próprio Apache Ranger para impor um controle refinado de acesso aos dados na Amazon. EMR Consulte [Integrar a Amazon EMR com o Apache Ranger](#) no Guia de EMR gerenciamento da Amazon.

- Políticas gerenciadas com escopo definido: para se alinhar às AWS melhores práticas, a Amazon EMR introduziu políticas EMR gerenciadas padrão com escopo v2 como substitutas das políticas que serão descontinuadas. Consulte as [políticas EMR gerenciadas da Amazon](#).
- Status de suporte do Instance Metadata Service (IMDS) V2: Para o Amazon EMR 6.2 ou posterior, os EMR componentes da Amazon são usados IMDSv2 para todas as chamadas. IMDS Para IMDS chamadas no código do seu aplicativo, você pode usar ambos IMDSv1 e IMDSv2, ou configurar o IMDS para uso somente IMDSv2 para aumentar a segurança. Se você desabilitar IMDSv1 em versões anteriores do Amazon EMR 6.x, isso causará uma falha na inicialização do cluster.

Alterações, melhorias e problemas resolvidos

- Esta é uma versão para corrigir problemas com o Amazon EMR Scaling quando ele falha na escalabilidade ou redução de escala de um cluster com sucesso ou causa falhas no aplicativo.
- Corrigido um problema em que as solicitações de escalabilidade falhavam em um cluster grande e altamente utilizado quando os daemons no cluster da EMR Amazon estavam executando atividades de verificação de integridade, como coletar o estado do nó e o estado do YARN nó. HDFS Isso estava acontecendo porque os daemons no cluster não conseguiam comunicar os dados do status de saúde de um nó aos componentes internos da Amazon. EMR
- Daemons EMR no cluster aprimorados para rastrear corretamente os estados dos nós quando os endereços IP são reutilizados para melhorar a confiabilidade durante as operações de escalabilidade.
- [SPARK-29683](#). Corrigido um problema em que ocorriam falhas de trabalho durante a redução da escala verticalmente do cluster, pois o Spark presumia que todos os nós disponíveis estavam na lista de negação.
- [YARN-9011](#). Corrigido um problema em que falhas de trabalho ocorriam devido a uma condição de corrida no YARN descomissionamento quando o cluster tentava aumentar ou diminuir a escala.
- Foi corrigido o problema com falhas de etapas ou tarefas durante a escalabilidade do cluster, garantindo que os estados dos nós fossem sempre consistentes entre os EMR daemons no cluster da Amazon e/. YARN HDFS
- Corrigido um problema em que operações de cluster, como redução de escala e envio de etapas, falhavam em EMR clusters da Amazon habilitados com a autenticação Kerberos. Isso ocorreu

porque o daemon no EMR cluster da Amazon não renovou o tíquete Kerberos, que é necessário para se comunicar de forma segura com/em execução no nó primário. HDFS YARN

- As EMR versões mais recentes da Amazon corrigem o problema com um limite menor de “Máximo de arquivos abertos” em relação aos mais antigos AL2 na AmazonEMR. As EMR versões 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 e posteriores agora incluem uma correção permanente com uma configuração mais alta de “Máximo de arquivos abertos”.
- O padrão do modo de explicação da SQL interface do Spark foi alterado de extended para formatted no [Spark 3.1](#). A Amazon o EMR reverteu extended para incluir informações do plano lógico na interface do usuário do SparkSQL. Isso pode ser revertido definindo `spark.sql.ui.explainMode` como formatted.
- As confirmações a seguir foram transferidas da principal ramificação do Spark.
 - [\[SPARK-34752\]](#) [\[BUILD\]](#) Bump Jetty para 9.4.37 para o endereço -2020-27223. CVE
 - [\[SPARK-34534\]](#) Corrige a blockIds ordem ao usar FetchShuffleBlocks para buscar blocos.
 - [\[SPARK-34681\]](#) [\[SQL\]](#) Corrige o erro da junção de hash embaralhada externa completa ao construir o lado esquerdo com condições diferentes.
 - [\[SPARK-34497\]](#) [\[SQL\]](#) Corrija os provedores de JDBC conexão integrados para restaurar as alterações no contexto JVM de segurança.
- [Para melhorar a interoperabilidade com o RAPIDS plug-in Nvidia Spark, foi adicionada uma solução alternativa para solucionar um problema que impede que a remoção dinâmica de partições seja acionada ao usar o Nvidia Spark RAPIDS com a execução adaptativa de consultas desativada, consulte o problema #1378 e o problema # #1386. RAPIDS RAPIDS](#) Para obter detalhes sobre a nova configuração `spark.sql.optimizer.dynamicPartitionPruning.enforceBroadcastReuse`, consulte o [RAPIDSproblema # #1386](#).
- O algoritmo padrão do confirmador de saída de arquivo foi alterado do algoritmo v2 para o algoritmo v1 no Spark 3.1 de código aberto. Para obter mais informações, consulte [Amazon EMR otimizando o desempenho do Spark - remoção dinâmica de partições](#).
- A Amazon EMR voltou ao algoritmo v2, o padrão usado em versões anteriores do Amazon EMR 6.x, para evitar a regressão de desempenho. Para restaurar o comportamento de código aberto do Spark 3.1, defina `spark.hadoop.mapreduce.fileoutputcommitter.algorithm.version` como 1. O Spark de código aberto fez essa alteração porque a confirmação de tarefas no algoritmo do confirmador de saída de arquivo v2 não é atômica, o que pode causar um problema de exatidão dos dados de saída em alguns casos. No entanto, a confirmação da tarefa no

algoritmo v1 também não é atômica. Em alguns cenários, a confirmação da tarefa inclui uma exclusão realizada antes de uma renomeação. Isso pode resultar em um problema silencioso de correção de dados.

- Corrigiu problemas de escalabilidade gerenciada em EMR versões anteriores da Amazon e fez melhorias para que as taxas de falha de aplicativos fossem significativamente reduzidas.
- Instalou o AWS Java SDK Bundle em cada novo cluster. Esse é um único jar contendo todos os serviços SDKs e suas dependências, em vez de jars de componentes individuais. Para obter mais informações, consulte [Java SDK Bundled Dependency](#).

Problemas conhecidos

- Para clusters de sub-rede privados Amazon EMR 6.3.0 e 6.2.0, você não pode acessar a interface web do Ganglia. Você receberá um erro de “acesso negado (403)”. Outros sites UIs, como Spark, Hue, Zeppelin JupyterHub, Livy e Tez, estão funcionando normalmente. O acesso à interface do usuário da Web do Ganglia em clusters de sub-redes públicas também está funcionando normalmente. Para resolver esse problema, reinicie o serviço httpd no nó primário com `sudo systemctl restart httpd`. Esse problema foi corrigido no Amazon EMR 6.4.0.
- Quando o AWS Glue Data Catalog está ativado, o uso do Spark para acessar um AWS Glue DB com localização de string nula URI pode falhar. Isso acontece com EMR versões anteriores da Amazon, mas SPARK -31709 (<https://issues.apache.org/jira/browse/SPARK-31709>) faz com que se aplique a mais casos. Por exemplo, ao criar uma tabela no AWS Glue DB padrão cuja localização URI é uma string nula, `spark.sql("CREATE TABLE mytest (key string) location '/table_path' ;")` falha com a mensagem “Não é possível criar um caminho a partir de uma string vazia”. Para contornar isso, defina manualmente a localização URI dos seus bancos de dados AWS Glue e crie tabelas dentro desses bancos de dados usando o Spark.
- Na Amazon EMR 6.3.0, o Presto SQL foi atualizado da versão 343 para a versão 350. Há duas alterações relacionadas à segurança provenientes do código aberto que se relaciona a essa alteração de versão. O controle de acesso ao catálogo baseado em arquivos é alterado de deny para allow quando as regras de propriedades de tabela, esquema ou sessão não são definidas. Além disso, o controle de acesso ao sistema baseado em arquivos é alterado para oferecer suporte a arquivos sem regras de catálogo definidas. Nesse caso, todo o acesso aos catálogos é permitido.

Para obter mais informações, consulte a [Versão 344 \(9 de outubro de 2020\)](#).

- Observe que o diretório de usuário do Hadoop (/home/hadoop) pode ser lido por todos. Ele tem permissões de diretório Unix 755 (drwxr-xr-x) para permitir acesso de leitura por estruturas como

o Hive. Você pode colocar arquivos em `/home/hadoop` e seus subdiretórios, mas esteja ciente das permissões desses diretórios para proteger informações confidenciais.

- Limite inferior de “Máximo de arquivos abertos” em versões mais antigas AL2 [corrigido em versões mais recentes]. EMRLançamentos da Amazon: `emr-5.30.x`, `emr-5.31.0`, `emr-5.32.0`, `emr-6.0.0`, `emr-6.1.0` e `emr-6.2.0` são baseados em versões mais antigas do Linux ofAmazon 2 (AL2, que têm uma configuração de limite inferior para “Máximo de arquivos abertos” quando os clusters da Amazon são criados com o padrão. EMR AMI As EMR versões `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` e posteriores incluem uma correção permanente com uma configuração maior de “Máximo de arquivos abertos”. Versões com o limite inferior de arquivos abertos causam o erro “Muitos arquivos abertos” ao ser enviado um trabalho do Spark. Nas versões afetadas, o padrão da Amazon AMI tem uma configuração EMR padrão de `ulimit` de 4096 para “Máximo de arquivos abertos”, que é inferior ao limite de 65536 arquivos no Linux 2. latestAmazon AMI A configuração inferior de `ulimit` para “Máximo de arquivos abertos” causa falhas em trabalhos do Spark quando o driver e o executor do Spark tentam abrir mais de 4096 arquivos. Para corrigir o problema, a Amazon EMR tem um script de ação de bootstrap (BA) que ajusta a configuração `ulimit` na criação do cluster.

Se você estiver usando uma EMR versão mais antiga da Amazon que não tem a solução permanente para esse problema, a solução alternativa a seguir permite que você defina explicitamente o `ulimit` do controlador de instância para um máximo de 65536 arquivos.

Defina explicitamente um `ulimit` na linha de comando

1. Edite `/etc/systemd/system/instance-controller.service` para adicionar os seguintes parâmetros à seção Serviço.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Reiniciar InstanceController


```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Defina um `ulimit` usando a ação de bootstrap (BA)

Você também pode usar um script de ação de bootstrap (BA) para configurar o `ulimit instance-controller` para 65536 arquivos na criação do 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

EMRclusters que executam Amazon Linux ou Amazon Linux 2 Amazon Machine Images (AMIs) usam o comportamento padrão do Amazon Linux e não baixam e instalam automaticamente atualizações importantes e críticas do kernel que exigem uma reinicialização. Esse é o mesmo comportamento de outras EC2 instâncias da Amazon que executam o Amazon Linux padrãoAMI. Se novas atualizações de software Amazon Linux que exigem uma reinicialização (como kernel e CUDA atualizações) ficarem disponíveis após a disponibilização de uma EMR versão da Amazon, as instâncias de EMR cluster que executam o padrão AMI não baixam e instalam automaticamente essas atualizações. NVIDIA Para obter atualizações do kernel, você pode [personalizar sua Amazon EMR AMI](#) para [usar o Amazon Linux AMI mais recente](#).

- Para usar as ações do Spark com o Apache Oozie, você deve adicionar a seguinte configuração ao seu arquivo `workflow.xml` do Oozie. Caso contrário, várias bibliotecas críticas, como a Hadoop e a Hadoop, EMRFS estarão ausentes do classpath dos executores do Spark que o Oozie lança.

```
<spark-opts>--conf spark.yarn.populateHadoopClasspath=true</spark-opts>
```

- Quando você usa o Spark com a formatação de localização de partições do Hive para ler dados no Amazon S3 e executa o Spark nas EMR versões 5.30.0 a 5.36.0 e 6.2.0 a 6.9.0 da Amazon, você pode encontrar um problema que impede seu cluster de ler os dados corretamente. Isso poderá acontecer se suas partições tiverem todas as características a seguir:
 - Duas ou mais partições são verificadas na mesma tabela.
 - Pelo menos um caminho de diretório de partição é um prefixo de pelo menos outro caminho de diretório de partição, por exemplo, `s3://bucket/table/p=a` é um prefixo de `s3://bucket/table/p=a b`.
 - O primeiro caractere que segue o prefixo no outro diretório de partição tem um valor UTF -8 menor que o / caractere (U+002F). Por exemplo, o caractere de espaço (U+0020) que ocorre entre a e b em `s3://bucket/table/p=a b` se enquadra nessa categoria. Observe que existem 14 outros caracteres que não são de controle: `!"#$%&'()*+,-.` Para obter mais informações, consulte a [tabela de codificação UTF -8 e caracteres Unicode](#).

Como solução alternativa para esse problema, defina a configuração `spark.sql.sources.fastS3PartitionDiscovery.enabled` como `false` na classificação `spark-defaults`.

Versões de componente 6.3.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion-amzn-EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK
emr-ddb	4.16.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	3.2.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.5.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-notebook-env	1.2.0	Ambiente Conda para bloco de anotações do emr, que inclui o jupyter enterprise gateway
emr-s3-dist-cp	2.18.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	2.1.0	EMRConector S3Select
emrfs	2.46.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.12.1	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
flink-jobmanager-config	1.12.1	Gerenciando recursos em EMR nós para o Apache JobManager Flink.

Componente	Version (Versão)	Descrição
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	3.2.1-amzn-3	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	3.2.1-amzn-3	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	3.2.1-amzn-3	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	3.2.1-amzn-3	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	3.2.1-amzn-3	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httpfs-server	3.2.1-amzn-3	HTTPendpoint para HDFS operações.

Componente	Version (Versão)	Descrição
hadoop-kms-server	3.2.1-amzn-3	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	3.2.1-amzn-3	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	3.2.1-amzn-3	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	3.2.1-amzn-3	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	3.2.1-amzn-3	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	2.2.6-amzn-1	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	2.2.6-amzn-1	Serviço para atender a uma ou mais HBase regiões.
hbase-client	2.2.6-amzn-1	HBasecliente de linha de comando.
hbase-rest-server	2.2.6-amzn-1	Serviço que fornece um RESTful HTTP endpoint paraHBase.

Componente	Version (Versão)	Descrição
hbase-thrift-server	2.2.6-amzn-1	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	3.1.2-amzn-4	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	3.1.2-amzn-4	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	3.1.2-amzn-4	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	3.1.2-amzn-4	O cliente da linha de comando do Hive.
hive-hbase	3.1.2-amzn-4	Cliente do Hive-hbase.
hive-metastore-server	3.1.2-amzn-4	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	3.1.2-amzn-4	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.7.0-amzn-0	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.

Componente	Version (Versão)	Descrição
hudi-presto	0.7.0-amzn-0	Biblioteca de pacotes para executar o Presto com o Hudi.
hudi-prestosql	0.7.0-amzn-0	Biblioteca de pacotes para executar o Presto SQL com o Hudi.
hudi-spark	0.7.0-amzn-0	Biblioteca de pacotes para executar o Spark com o Hudi.
hue-server	4.9.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	1.2.2	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.7.0-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mxnet	1.7.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mariadb-server	5.5.68+	Servidor de banco de dados MariaDB.
nvidia-cuda	10.1.243	Drivers NVIDIA e toolkit CUDA
oozie-client	5.2.1	O cliente da linha de comando do Oozie.

Componente	Version (Versão)	Descrição
oozie-server	5.2.1	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	4.5.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	5.0.0- HBase -2,0	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	5.0.0- HBase -2,0	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.245.1-amzn-0	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.245.1-amzn-0	O serviço que executa partes de uma consulta.
presto-client	0.245.1-amzn-0	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
prestosql-coordinator	350	O serviço que aceita consultas e gerencia a execução de consultas entre os prestosql-workers.
prestosql-worker	350	O serviço que executa partes de uma consulta.

Componente	Version (Versão)	Descrição
prestosql-client	350	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	4.0.2	O projeto R para computação estatística
ranger-kms-server	2.0.0	Sistema de gerenciamento de chaves Apache Ranger
spark-client	3.1.1-amzn-0	Os clientes da linha de comando do Spark.
spark-history-server	3.1.1-amzn-0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	3.1.1-amzn-0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	3.1.1-amzn-0	Bibliotecas do Apache Spark necessárias para escravos. YARN
spark-rapids	0.4.1	RAPIDSPlugin Nvidia Spark que acelera o Apache Spark com o. GPUs

Componente	Version (Versão)	Descrição
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	2.4.1	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.2	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.41+	HTTPServidor Apache.
zeppelin-server	0.9.0	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.14	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.14	ZooKeeper cliente de linha de comando.

Classificações de configuração 6.3.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Ocorrem ações de reconfiguração quando você especifica uma configuração para grupos de instâncias em um cluster em execução. A Amazon EMR só inicia ações de reconfiguração para as

classificações que você modifica. Para ter mais informações, consulte [Reconfigurar um grupo de instâncias de um cluster em execução](#).

Classificações do emr-6.3.0

Classificações	Descrição	Ações de reconfiguração
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.	Restarts the Resource Manager service.
container-executor	Altere os valores no arquivo container-executor.cfg YARN do Hadoop.	Not available.
container-log4j	Altere os valores no arquivo container-log4j.properties YARN do Hadoop.	Not available.
core-site	Alterar os valores no arquivo core-site.xml do 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	Alterar as configurações relacionadas ao docker.	Not available.
emrfs-site	Altere EMRFS as configurações.	Restarts the Hadoop HDFS services Namenode,

Classificações	Descrição	Ações de reconfiguração
		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	Alterar as configurações do <code>flink-conf.yaml</code> .	Restarts Flink history server.
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.	Restarts Flink history server.
flink-log4j-session	Alterar as configurações de <code>log4j-yarn-session.properties</code> no Flink para sessão do Kubernetes/Yarn.	Restarts Flink history server.
flink-log4j-cli	Alterar as configurações de <code>log4j-cli.properties</code> no Flink.	Restarts Flink history server.

Classificações	Descrição	Ações de reconfiguração
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do 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	Alterar os valores no arquivo log4j.properties do 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	Alterar a configuração do servidor SSL no Hadoop	Not available.
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop	Not available.

Classificações	Descrição	Ações de reconfiguração
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Mude os valores no HBase ambiente.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.	Not available.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.	This classification should not be reconfigured.
hdfs-env	Mude os valores no HDFS ambiente.	Restarts Hadoop HDFS services Namenode, Datanode, and ZKFC.

Classificações	Descrição	Ações de reconfiguração
hdfs-site	Altere os valores no HDFS hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Mude os valores no HCatalog ambiente.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Configurações EMR selecionadas pela Amazon para o Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.	Not available.

Classificações	Descrição	Ações de reconfiguração
hive-parquet-logging	Alterar os valores no arquivo <code>parquet-logging.properties</code> do Hive.	Not available.
hive-env	Alterar os valores no ambiente do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive <code>schemaTool CLI</code> commands to verify <code>hive-metastore</code> .
hive-exec-log4j2	Altere os valores no arquivo <code>hive-exec-log 4j2.properties</code> do Hive.	Not available.
hive-llap-daemon-log4j2	Altere os valores no arquivo <code>llap-daemon-log 4j2.properties</code> do Hive.	Not available.
hive-log4j2	Alterar os valores no arquivo <code>hive-log4j2.properties</code> do Hive.	Not available.
hive-site	Alterar os valores no arquivo <code>hive-site.xml</code> do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive <code>schemaTool CLI</code> commands to verify <code>hive-metastore</code> . Also restarts Oozie and Zeppelin.
hiveserver2-site	Alterar os valores no arquivo <code>hiveserver2-site.xml</code> do Hive Server2.	Not available.
hue-ini	Alterar os valores no arquivo <code>ini</code> do Hue	Restarts Hue. Also activates Hue <code>config override CLI</code> commands to pick up new configurations.

Classificações	Descrição	Ações de reconfiguração
httpfs-env	Mude os valores no HTTPFS ambiente.	Restarts Hadoop Httpfs service.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.	Not available.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.	Not available.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Altere os valores no ambiente do Hudi.	Not available.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.	Not available.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.	Not available.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.	Not available.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.	Not available.

Classificações	Descrição	Ações de reconfiguração
livy-conf	Alterar os valores no arquivo livy.conf do Livy.	Restarts Livy Server.
livy-env	Alterar os valores no ambiente do Livy.	Restarts Livy Server.
livy-log4j	Alterar as configurações de log4j.properties no Livy.	Restarts Livy Server.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Alterar os valores no ambiente do Oozie.	Restarts Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.	Restarts Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.	Not available.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.	Not available.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.	Restarts Phoenix-QueryServer.

Classificações	Descrição	Ações de reconfiguração
phoenix-metrics	Alterar os valores no arquivo <code>hadoop-metrics2-phoenix.properties</code> do Phoenix.	Not available.
pig-env	Alterar os valores no ambiente do Pig.	Not available.
pig-properties	Alterar os valores no arquivo <code>pig.properties</code> do Pig.	Restarts Oozie.
pig-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Pig.	Not available.
presto-log	Alterar os valores no arquivo <code>log.properties</code> do Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Alterar os valores no arquivo <code>config.properties</code> do Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Altere os valores no arquivo <code>password-authenticator.properties</code> do Presto.	Not available.
presto-env	Altere os valores no arquivo <code>presto-env.sh</code> do Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Altere os valores no arquivo <code>node.properties</code> do Presto.	Not available.
presto-connector-blackhole	Alterar os valores no arquivo <code>blackhole.properties</code> do Presto.	Not available.
presto-connector-cassandra	Alterar os valores no arquivo <code>cassandra.properties</code> do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.	Not available.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.	Not available.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.	Not available.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.	Not available.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.	Not available.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.	Not available.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.	Not available.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.	Not available.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.	Not available.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.	Not available.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.	Not available.
prestosql-log	Alterar os valores no arquivo log.properties do Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-config	Alterar os valores no arquivo config.properties do Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-env	Altere os valores no arquivo presto-env.sh do Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-node	Altere os valores no arquivo SQL node.properties do Presto.	Not available.
prestosql-connector-blackhole	Altere os valores no arquivo SQL blackhole.properties do Presto.	Not available.
prestosql-connector-cassandra	Altere os valores no arquivo SQL cassandra.properties do Presto.	Not available.
prestosql-connector-hive	Altere os valores no arquivo SQL hive.properties do Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-connector-jmx	Altere os valores no arquivo SQL jmx.properties do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
prestosql-connector-kafka	Altere os valores no arquivo SQL kafka.properties do Presto.	Not available.
prestosql-connector-localfile	Altere os valores no arquivo SQL localfile.properties do Presto.	Not available.
prestosql-connector-memory	Altere os valores no arquivo SQL memory.properties do Presto.	Not available.
prestosql-connector-mongodb	Altere os valores no arquivo SQL mongodb.properties do Presto.	Not available.
prestosql-connector-mysql	Altere os valores no arquivo SQL mysql.properties do Presto.	Not available.
prestosql-connector-postgresql	Altere os valores no arquivo SQL postgresql.properties do Presto.	Not available.
prestosql-connector-raptor	Altere os valores no arquivo SQL raptor.properties do Presto.	Not available.
prestosql-connector-redis	Altere os valores no arquivo SQL redis.properties do Presto.	Not available.
prestosql-connector-redshift	Altere os valores no arquivo SQL redshift.properties do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
prestosql-connector-tpch	Altere os valores no arquivo SQL tpch.properties do Presto.	Not available.
prestosql-connector-tpcds	Altere os valores no arquivo SQL tpcds.properties do Presto.	Not available.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Altere os valores no arquivo kms-log4j.properties do Ranger. KMS	Not available.
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.	Not available.
spark	Configurações EMR seleciona das pela Amazon para o Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Alterar os valores no ambiente do Spark.	Restarts Spark history server and Spark thrift server.

Classificações	Descrição	Ações de reconfiguração
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.	Not available.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Alterar os valores no ambiente do Sqoop.	Not available.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.	Not available.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.	Not available.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.	Restart Oozie and HiveServer2.
yarn-env	Mude os valores no YARN ambiente.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.

Classificações	Descrição	Ações de reconfiguração
zeppelin-env	Alterar os valores no ambiente do Zeppelin.	Restarts Zeppelin.
zeppelin-site	Alterar as definições da configuração em zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.	Restarts Zookeeper server.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.	Restarts Zookeeper server.

Amazon EMR versão 6.2.1

Versões 6.2.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [PrestoSQL](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-6.2.1	emr-6.2.0	emr-6.1.1	emr-6.1.0
AWS SDK para Java	1.11.880	1.11.880	1.11.828	1.11.828
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12.10	2.12.10	2.12.10	2.12.10
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

Notas da versão 6.2.1

Esta é uma versão para corrigir problemas com o Amazon EMR Scaling quando ele falha na escalabilidade ou redução de escala de um cluster com sucesso ou causa falhas no aplicativo.

Alterações, melhorias e problemas resolvidos

- Corrigido um problema em que as solicitações de escalabilidade falhavam em um cluster grande e altamente utilizado quando os daemons no cluster da EMR Amazon estavam executando atividades de verificação de integridade, como coletar o estado do nó e o estado do YARN nó. HDFS Isso estava acontecendo porque os daemons no cluster não conseguiam comunicar os dados do status de saúde de um nó aos componentes internos da Amazon. EMR

- Daemons EMR no cluster aprimorados para rastrear corretamente os estados dos nós quando os endereços IP são reutilizados para melhorar a confiabilidade durante as operações de escalabilidade.
- [SPARK-29683](#). Corrigido um problema em que ocorriam falhas de trabalho durante a redução da escala verticalmente do cluster, pois o Spark presumia que todos os nós disponíveis estavam na lista de negação.
- [YARN-9011](#). Corrigido um problema em que falhas de trabalho ocorriam devido a uma condição de corrida no YARN descomissionamento quando o cluster tentava aumentar ou diminuir a escala.
- Foi corrigido o problema com falhas de etapas ou tarefas durante a escalabilidade do cluster, garantindo que os estados dos nós fossem sempre consistentes entre os EMR daemons no cluster da Amazon e/. YARN HDFS
- Corrigido um problema em que operações de cluster, como redução de escala e envio de etapas, falhavam em EMR clusters da Amazon habilitados com a autenticação Kerberos. Isso ocorreu porque o daemon no EMR cluster da Amazon não renovou o tíquete Kerberos, que é necessário para se comunicar de forma segura com/em execução no nó primário. HDFS YARN
- As EMR versões mais recentes da Amazon corrigem o problema com um limite menor de “Máximo de arquivos abertos” em relação aos mais antigos AL2 na AmazonEMR. As EMR versões 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 e posteriores agora incluem uma correção permanente com uma configuração mais alta de “Máximo de arquivos abertos”.
- HTTPS agora está habilitado por padrão para os repositórios Amazon Linux. Se você estiver usando uma VPCE política do Amazon S3 para restringir o acesso a buckets específicos, deverá adicionar o novo bucket Amazon Linux ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` à sua política (`$region` substitua pela região onde está o endpoint). Para obter mais informações, consulte esse tópico nos fóruns de AWS discussão. [Anúncio: o Amazon Linux 2 agora oferece suporte à capacidade de uso HTTPS ao se conectar a repositórios de pacotes.](#)

Problemas conhecidos

- Quando você usa o Spark com a formatação de localização de partições do Hive para ler dados no Amazon S3 e executa o Spark nas EMR versões 5.30.0 a 5.36.0 e 6.2.0 a 6.9.0 da Amazon, você pode encontrar um problema que impede seu cluster de ler os dados corretamente. Isso poderá acontecer se suas partições tiverem todas as características a seguir:
 - Duas ou mais partições são verificadas na mesma tabela.

- Pelo menos um caminho de diretório de partição é um prefixo de pelo menos outro caminho de diretório de partição, por exemplo, `s3://bucket/table/p=a` é um prefixo de `s3://bucket/table/p=a b`.
- O primeiro caractere que segue o prefixo no outro diretório de partição tem um valor UTF -8 menor que o / caractere (U+002F). Por exemplo, o caractere de espaço (U+0020) que ocorre entre a e b em `s3://bucket/table/p=a b` se enquadra nessa categoria. Observe que existem 14 outros caracteres que não são de controle: `!"#$%&'()*+,-.` Para obter mais informações, consulte a [tabela de codificação UTF -8 e caracteres Unicode](#).

Como solução alternativa para esse problema, defina a configuração `spark.sql.sources.fastS3PartitionDiscovery.enabled` como `false` na classificação `spark-defaults`.

Versões de componente 6.2.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.16.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
emr-goodies	3.1.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.5.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-notebook-env	1.0.0	Ambiente Conda para bloco de anotações do emr, que inclui o jupyter enterprise gateway
emr-s3-dist-cp	2.16.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	2.0.0	EMRConector S3Select
emrfs	2.44.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.11.2	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
flink-jobmanager-config	1.11.2	Gerenciando recursos em EMR nós para o Apache JobManager Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.

Componente	Version (Versão)	Descrição
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	3.2.1-amzn-2.1	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	3.2.1-amzn-2.1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	3.2.1-amzn-2.1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	3.2.1-amzn-2.1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	3.2.1-amzn-2.1	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httpfs-server	3.2.1-amzn-2.1	HTTPendpoint para HDFS operações.
hadoop-kms-server	3.2.1-amzn-2.1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API

Componente	Version (Versão)	Descrição
hadoop-mapred	3.2.1-amzn-2.1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	3.2.1-amzn-2.1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	3.2.1-amzn-2.1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	3.2.1-amzn-2.1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	2.2.6-amzn-0	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	2.2.6-amzn-0	Serviço para atender a uma ou mais HBase regiões.
hbase-client	2.2.6-amzn-0	HBasecliente de linha de comando.
hbase-rest-server	2.2.6-amzn-0	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	2.2.6-amzn-0	Serviço que fornece um endpoint Thrift para. HBase

Componente	Version (Versão)	Descrição
hcatalog-client	3.1.2-amzn-3	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	3.1.2-amzn-3	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	3.1.2-amzn-3	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	3.1.2-amzn-3	O cliente da linha de comando do Hive.
hive-hbase	3.1.2-amzn-3	Cliente do Hive-hbase.
hive-metastore-server	3.1.2-amzn-3	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	3.1.2-amzn-3	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.6.0-amzn-1	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.
hudi-presto	0.6.0-amzn-1	Biblioteca de pacotes para executar o Presto com o Hudi.

Componente	Version (Versão)	Descrição
hudi-prestosql	0.6.0-amzn-1	Biblioteca de pacotes para executar o Presto SQL com o Hudi.
hudi-spark	0.6.0-amzn-1	Biblioteca de pacotes para executar o Spark com o Hudi.
hue-server	4.8.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	1.1.0	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.7.0-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mxnet	1.7.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mariadb-server	5.5.64+	Servidor de banco de dados MariaDB.
nvidia-cuda	10.1.243	Drivers NVIDIA e toolkit CUDA
oozie-client	5.2.0	O cliente da linha de comando do Oozie.
oozie-server	5.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.

Componente	Version (Versão)	Descrição
opencv	4.4.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	5.0.0- HBase -2,0	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	5.0.0- HBase -2,0	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.238.3-amzn-1	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.238.3-amzn-1	O serviço que executa partes de uma consulta.
presto-client	0.238.3-amzn-1	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
prestosql-coordinator	343	O serviço que aceita consultas e gerencia a execução de consultas entre os prestosql-workers.
prestosql-worker	343	O serviço que executa partes de uma consulta.

Componente	Version (Versão)	Descrição
prestosql-client	343	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.3	O projeto R para computação estatística
ranger-kms-server	2.0.0	Sistema de gerenciamento de chaves Apache Ranger
spark-client	3.0.1-amzn-0.1	Os clientes da linha de comando do Spark.
spark-history-server	3.0.1-amzn-0.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	3.0.1-amzn-0.1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	3.0.1-amzn-0.1	Bibliotecas do Apache Spark necessárias para escravos. YARN
spark-rapids	0.2.0	RAPIDSPlugin Nvidia Spark que acelera o Apache Spark com o. GPUs

Componente	Version (Versão)	Descrição
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	2.3.1	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.2	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.41+	HTTPServidor Apache.
zeppelin-server	0.9.0-preview1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.14	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.14	ZooKeeper cliente de linha de comando.

Classificações de configuração 6.2.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Ocorrem ações de reconfiguração quando você especifica uma configuração para grupos de instâncias em um cluster em execução. A Amazon EMR só inicia ações de reconfiguração para as

classificações que você modifica. Para ter mais informações, consulte [Reconfigurar um grupo de instâncias de um cluster em execução](#).

Classificações do emr-6.2.1

Classificações	Descrição	Ações de reconfiguração
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.	Restarts the ResourceManager service.
container-executor	Altere os valores no arquivo container-executor.cfg YARN do Hadoop.	Not available.
container-log4j	Altere os valores no arquivo container-log4j.properties YARN do Hadoop.	Not available.
core-site	Alterar os valores no arquivo core-site.xml do 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	Alterar as configurações relacionadas ao docker.	Not available.
emrfs-site	Altere EMRFS as configurações.	Restarts the Hadoop HDFS services Namenode,

Classificações	Descrição	Ações de reconfiguração
		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	Alterar as configurações do <code>flink-conf.yaml</code> .	Not available.
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.	Not available.
flink-log4j-yarn-session	Altere as configurações do <code>Flink log4j-yarn-session.properties</code> .	Not available.
flink-log4j-cli	Alterar as configurações de <code>log4j-cli.properties</code> no Flink.	Not available.

Classificações	Descrição	Ações de reconfiguração
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do 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	Alterar os valores no arquivo log4j.properties do 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	Alterar a configuração do servidor SSL no Hadoop	Not available.
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop	Not available.

Classificações	Descrição	Ações de reconfiguração
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Mude os valores no HBase ambiente.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.	Not available.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.	This classification should not be reconfigured.
hdfs-env	Mude os valores no HDFS ambiente.	Restarts Hadoop HDFS ZKFC.

Classificações	Descrição	Ações de reconfiguração
hdfs-site	Altere os valores no HDFS hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Mude os valores no HCatalog ambiente.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Configurações EMR selecionadas pela Amazon para o Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.	Not available.

Classificações	Descrição	Ações de reconfiguração
hive-parquet-logging	Alterar os valores no arquivo <code>parquet-logging.properties</code> do Hive.	Not available.
hive-env	Alterar os valores no ambiente do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive <code>schemaTool CLI</code> commands to verify <code>hive-metastore</code> .
hive-exec-log4j2	Altere os valores no arquivo <code>hive-exec-log 4j2.properties</code> do Hive.	Not available.
hive-llap-daemon-log4j2	Altere os valores no arquivo <code>llap-daemon-log 4j2.properties</code> do Hive.	Not available.
hive-log4j2	Alterar os valores no arquivo <code>hive-log4j2.properties</code> do Hive.	Not available.
hive-site	Alterar os valores no arquivo <code>hive-site.xml</code> do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive <code>schemaTool CLI</code> commands to verify <code>hive-metastore</code> . Also restarts Oozie and Zeppelin.
hiveserver2-site	Alterar os valores no arquivo <code>hiveserver2-site.xml</code> do Hive Server2.	Not available.
hue-ini	Alterar os valores no arquivo <code>ini</code> do Hue	Restarts Hue. Also activates Hue <code>config override CLI</code> commands to pick up new configurations.

Classificações	Descrição	Ações de reconfiguração
httpfs-env	Mude os valores no HTTPFS ambiente.	Restarts Hadoop Httpfs service.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.	Not available.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.	Not available.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Altere os valores no ambiente do Hudi.	Not available.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.	Not available.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.	Not available.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.	Not available.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.	Not available.

Classificações	Descrição	Ações de reconfiguração
livy-conf	Alterar os valores no arquivo livy.conf do Livy.	Restarts Livy Server.
livy-env	Alterar os valores no ambiente do Livy.	Restarts Livy Server.
livy-log4j	Alterar as configurações de log4j.properties no Livy.	Restarts Livy Server.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Alterar os valores no ambiente do Oozie.	Restarts Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.	Restarts Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.	Not available.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.	Not available.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.	Restarts Phoenix-QueryServer.

Classificações	Descrição	Ações de reconfiguração
phoenix-metrics	Alterar os valores no arquivo <code>hadoop-metrics2-phoenix.properties</code> do Phoenix.	Not available.
pig-env	Alterar os valores no ambiente do Pig.	Not available.
pig-properties	Alterar os valores no arquivo <code>pig.properties</code> do Pig.	Restarts Oozie.
pig-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Pig.	Not available.
presto-log	Alterar os valores no arquivo <code>log.properties</code> do Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Alterar os valores no arquivo <code>config.properties</code> do Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Altere os valores no arquivo <code>password-authenticator.properties</code> do Presto.	Not available.
presto-env	Altere os valores no arquivo <code>presto-env.sh</code> do Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Altere os valores no arquivo <code>node.properties</code> do Presto.	Not available.
presto-connector-blackhole	Alterar os valores no arquivo <code>blackhole.properties</code> do Presto.	Not available.
presto-connector-cassandra	Alterar os valores no arquivo <code>cassandra.properties</code> do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.	Not available.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.	Not available.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.	Not available.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.	Not available.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.	Not available.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.	Not available.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.	Not available.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.	Not available.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.	Not available.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.	Not available.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.	Not available.
prestosql-log	Alterar os valores no arquivo log.properties do Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-config	Alterar os valores no arquivo config.properties do Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-env	Altere os valores no arquivo presto-env.sh do Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-node	Altere os valores no arquivo SQL node.properties do Presto.	Not available.
prestosql-connector-blackhole	Altere os valores no arquivo SQL blackhole.properties do Presto.	Not available.
prestosql-connector-cassandra	Altere os valores no arquivo SQL cassandra.properties do Presto.	Not available.
prestosql-connector-hive	Altere os valores no arquivo SQL hive.properties do Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-connector-jmx	Altere os valores no arquivo SQL jmx.properties do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
prestosql-connector-kafka	Altere os valores no arquivo SQL kafka.properties do Presto.	Not available.
prestosql-connector-localfile	Altere os valores no arquivo SQL localfile.properties do Presto.	Not available.
prestosql-connector-memory	Altere os valores no arquivo SQL memory.properties do Presto.	Not available.
prestosql-connector-mongodb	Altere os valores no arquivo SQL mongodb.properties do Presto.	Not available.
prestosql-connector-mysql	Altere os valores no arquivo SQL mysql.properties do Presto.	Not available.
prestosql-connector-postgresql	Altere os valores no arquivo SQL postgresql.properties do Presto.	Not available.
prestosql-connector-raptor	Altere os valores no arquivo SQL raptor.properties do Presto.	Not available.
prestosql-connector-redis	Altere os valores no arquivo SQL redis.properties do Presto.	Not available.
prestosql-connector-redshift	Altere os valores no arquivo SQL redshift.properties do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
prestosql-connector-tpch	Altere os valores no arquivo SQL tpch.properties do Presto.	Not available.
prestosql-connector-tpcds	Altere os valores no arquivo SQL tpcds.properties do Presto.	Not available.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Altere os valores no arquivo kms-log4j.properties do Ranger. KMS	Not available.
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.	Not available.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Alterar os valores no ambiente do Spark.	Restarts Spark history server and Spark thrift server.

Classificações	Descrição	Ações de reconfiguração
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.	Not available.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Alterar os valores no ambiente do Sqoop.	Not available.
sqoop-oracoop-site	Altere os valores no arquivo oracoop-site.xml OraOop do Sqoop.	Not available.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.	Not available.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.	Restart Oozie.
yarn-env	Mude os valores no YARN ambiente.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.

Classificações	Descrição	Ações de reconfiguração
zeppelin-env	Alterar os valores no ambiente do Zeppelin.	Restarts Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.	Restarts Zookeeper server.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.	Restarts Zookeeper server.

Amazon EMR versão 6.2.0

Versões 6.2.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [PrestoSQL](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-6.2.0	emr-6.1.1	emr-6.1.0	emr-6.0.1
AWS SDK para Java	1.11.880	1.11.828	1.11.828	1.11.711

	emr-6.2.0	emr-6.1.1	emr-6.1.0	emr-6.0.1
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12.10	2.12.10	2.12.10	2.12.10
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

	emr-6.2.0	emr-6.1.1	emr-6.1.0	emr-6.0.1
Phoenix	5.0.0	5.0.0	5.0.0	5.0.0
Pig	0.17.0	0.17.0	0.17.0	-
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

Notas da versão 6.2.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 6.2.0 da Amazon. As alterações são referentes à versão 6.1.0.

Data da versão inicial: 9 de dezembro de 2020

Data da última atualização: 4 de outubro de 2021

Aplicações compatíveis

- AWS SDK for Java versão 1.11.828
- emr-record-server versão 1.7.0
- Flink versão 1.11.2
- Ganglia versão 3.7.2
- Hadoop versão 3.2.1-amzn-1

- HBase versão 2.2.6-amzn-0
- HBase-operator-tools 1.0.0
- HCatalog versão 3.1.2-amzn-0
- Hive versão 3.1.2-amzn-3
- Hudi versão 0.6.0-amzn-1
- Hue versão 4.8.0
- JupyterHub versão 1.1.0
- Livy versão 0.7.0
- MXNet versão 1.7.0
- Oozie versão 5.2.0
- Phoenix versão 5.0.0
- Pig versão 0.17.0
- Presto versão 0.238.3-amzn-1
- Presto SQL versão 343
- Spark versão 3.0.1-amzn-0
- spark-rapids 0.2.0
- TensorFlow versão 2.3.1
- Zeppelin versão 0.9.0-preview1
- Zookeeper versão 3.4.14
- Conectores e drivers: DynamoDB Connector 4.16.0

Novos atributos

- HBase: a renomeação foi removida na fase de confirmação e o HFile rastreamento persistente foi adicionado. Consulte [HFileRastreamento persistente](#) no Guia EMR de lançamento da Amazon.
- HBase: Backported [Crie uma configuração que força o armazenamento em cache dos blocos na compactação](#).
- PrestoDB: melhorias na remoção dinâmica de partições. O Join Reorder baseado em regras funciona em dados não particionados.
- Políticas gerenciadas com escopo definido: para se alinhar às AWS melhores práticas, a Amazon EMR introduziu políticas EMR gerenciadas padrão com escopo v2 como substitutas das políticas que serão descontinuadas. Consulte as [políticas EMR gerenciadas da Amazon](#).

- Status de suporte do Instance Metadata Service (IMDS) V2: Para o Amazon EMR 6.2 ou posterior, os EMR componentes da Amazon são usados IMDSv2 para todas as chamadas. IMDS Para IMDS chamadas no código do seu aplicativo, você pode usar ambos IMDSv1 e IMDSv2, ou configurar o IMDS para uso somente IMDSv2 para aumentar a segurança. Se você desabilitar IMDSv1 em versões anteriores do Amazon EMR 6.x, isso causará uma falha na inicialização do cluster.

Alterações, melhorias e problemas resolvidos

- Esta é uma versão para corrigir problemas com o Amazon EMR Scaling quando ele falha na escalabilidade ou redução de escala de um cluster com sucesso ou causa falhas no aplicativo.
- Corrigido um problema em que as solicitações de escalabilidade falhavam em um cluster grande e altamente utilizado quando os daemons no cluster da EMR Amazon estavam executando atividades de verificação de integridade, como coletar o estado do nó e o estado do YARN nó. HDFS Isso estava acontecendo porque os daemons no cluster não conseguiam comunicar os dados do status de saúde de um nó aos componentes internos da Amazon. EMR
- Daemons EMR no cluster aprimorados para rastrear corretamente os estados dos nós quando os endereços IP são reutilizados para melhorar a confiabilidade durante as operações de escalabilidade.
- [SPARK-29683](#). Corrigido um problema em que ocorriam falhas de trabalho durante a redução da escala verticalmente do cluster, pois o Spark presumia que todos os nós disponíveis estavam na lista de negação.
- [YARN-9011](#). Corrigido um problema em que falhas de trabalho ocorriam devido a uma condição de corrida no YARN descomissionamento quando o cluster tentava aumentar ou diminuir a escala.
- Foi corrigido o problema com falhas de etapas ou tarefas durante a escalabilidade do cluster, garantindo que os estados dos nós fossem sempre consistentes entre os EMR daemons no cluster da Amazon e/. YARN HDFS
- Corrigido um problema em que operações de cluster, como redução de escala e envio de etapas, falhavam em EMR clusters da Amazon habilitados com a autenticação Kerberos. Isso ocorreu porque o daemon no EMR cluster da Amazon não renovou o tíquete Kerberos, que é necessário para se comunicar de forma segura com/em execução no nó primário. HDFS YARN
- As EMR versões mais recentes da Amazon corrigem o problema com um limite menor de “Máximo de arquivos abertos” em relação aos mais antigos AL2 na AmazonEMR. As EMR versões 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 e posteriores agora incluem uma correção permanente com uma configuração mais alta de “Máximo de arquivos abertos”.
- Spark: melhorias de performance no runtime do Spark.

Problemas conhecidos

- O Amazon EMR 6.2 tem permissões incorretas definidas no arquivo `/etc/cron.d/libinstance-controller-java` na versão 6.2.0. EMR As permissões no arquivo são 645 (-rw-r--r-x), quando deveriam ser 644 (-rw-r--r--). Como resultado, a EMR versão 6.2 da Amazon não registra registros do estado da instância e o diretório `/emr/instance-logs` está vazio. Esse problema foi corrigido no Amazon EMR 6.3.0 e versões posteriores.

Para contornar esse problema, execute o script a seguir como uma ação de bootstrap na inicialização do cluster.

```
#!/bin/bash
sudo chmod 644 /etc/cron.d/libinstance-controller-java
```

- Para clusters de sub-rede privados Amazon EMR 6.2.0 e 6.3.0, você não pode acessar a interface web do Ganglia. Você receberá um erro de “acesso negado (403)”. Outros sites UIs, como Spark, Hue, Zeppelin JupyterHub, Livy e Tez, estão funcionando normalmente. O acesso à interface do usuário da Web do Ganglia em clusters de sub-redes públicas também está funcionando normalmente. Para resolver esse problema, reinicie o serviço `httpd` no nó primário com `sudo systemctl restart httpd`. Esse problema foi corrigido no Amazon EMR 6.4.0.
- Há um problema no Amazon EMR 6.2.0 em que o `httpd` falha continuamente, fazendo com que o Ganglia fique indisponível. Você recebe a mensagem de erro “cannot connect to the server”. Para corrigir um cluster que já está em execução com esse problema, SSH acesse o nó primário do cluster e adicione a linha `Listen 80` ao arquivo `httpd.conf` localizado em `/etc/httpd/conf/httpd.conf`. Esse problema foi corrigido no Amazon EMR 6.3.0.
- HTTPD falha nos clusters EMR 6.2.0 quando você usa uma configuração de segurança. Isso faz com que a interface de usuário da aplicação Web do Ganglia fique indisponível. Para acessar a interface de usuário da aplicação Web do Ganglia, adicione `Listen 80` ao arquivo `/etc/httpd/conf/httpd.conf` no nó primário do cluster. Para obter informações sobre como se conectar ao seu cluster, consulte [Conectar-se ao nó primário usando SSH](#).

EMROs notebooks também não conseguem estabelecer uma conexão com os clusters EMR 6.2.0 quando você usa uma configuração de segurança. O caderno não conseguirá listar os kernels e enviar trabalhos do Spark. Em EMR vez disso, recomendamos que você use EMR Notebooks com outra versão da Amazon.

- Limite inferior de “Máximo de arquivos abertos” em versões mais antigas AL2 [corrigido em versões mais recentes]. EMRLançamentos da Amazon: `emr-5.30.x`, `emr-5.31.0`, `emr-5.32.0`,

emr-6.0.0, emr-6.1.0 e emr-6.2.0 são baseados em versões mais antigas do Linux of Amazon 2 (AL2), que têm uma configuração de limite inferior para “Máximo de arquivos abertos” quando os clusters da Amazon são criados com o padrão. EMR AMI As EMR versões 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 e posteriores incluem uma correção permanente com uma configuração maior de “Máximo de arquivos abertos”. Versões com o limite inferior de arquivos abertos causam o erro “Muitos arquivos abertos” ao ser enviado um trabalho do Spark. Nas versões afetadas, o padrão da Amazon AMI tem uma configuração EMR padrão de ulimit de 4096 para “Máximo de arquivos abertos”, que é inferior ao limite de 65536 arquivos no Linux 2. latest Amazon AMI A configuração inferior de ulimit para “Máximo de arquivos abertos” causa falhas em trabalhos do Spark quando o driver e o executor do Spark tentam abrir mais de 4096 arquivos. Para corrigir o problema, a Amazon EMR tem um script de ação de bootstrap (BA) que ajusta a configuração ulimit na criação do cluster.

Se você estiver usando uma EMR versão mais antiga da Amazon que não tem a solução permanente para esse problema, a solução alternativa a seguir permite que você defina explicitamente o ulimit do controlador de instância para um máximo de 65536 arquivos.

Defina explicitamente um ulimit na linha de comando

1. Edite `/etc/systemd/system/instance-controller.service` para adicionar os seguintes parâmetros à seção Serviço.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Reiniciar InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Defina um ulimit usando a ação de bootstrap (BA)

Você também pode usar um script de ação de bootstrap (BA) para configurar o ulimit instance-controller para 65536 arquivos na criação do 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**

O Amazon EMR 6.1.0 e 6.2.0 incluem um problema de desempenho que pode afetar criticamente todas as operações de inserção, atualização e exclusão do Hudi. Se você planeja usar o Hudi com o Amazon EMR 6.1.0 ou 6.2.0, entre em contato com o AWS suporte para obter um Hudi corrigido. RPM

• **⚠ Important**

EMRclusters que executam Amazon Linux ou Amazon Linux 2 Amazon Machine Images (AMIs) usam o comportamento padrão do Amazon Linux e não baixam e instalam automaticamente atualizações importantes e críticas do kernel que exigem uma reinicialização. Esse é o mesmo comportamento de outras EC2 instâncias da Amazon que executam o Amazon Linux padrãoAMI. Se novas atualizações de software Amazon Linux que exigem uma reinicialização (como kernel e CUDA atualizações) ficarem disponíveis após a disponibilização de uma EMR versão da Amazon, as instâncias de EMR cluster que executam o padrão AMI não baixam e instalam automaticamente essas atualizações. NVIDIA Para obter atualizações do kernel, você pode [personalizar sua Amazon EMR AMI](#) para [usar o Amazon Linux AMI mais recente](#).

- Os artefatos do Amazon EMR 6.2.0 Maven não são publicados. Eles serão publicados com uma futura versão da AmazonEMR.

- O HFile rastreamento persistente usando a tabela do sistema HBase storefile não oferece suporte ao recurso de replicação da HBase região. Para obter mais informações sobre a replicação HBase da região, consulte Leituras de alta disponibilidade [consistentes com o cronograma](#).
- Diferenças entre as versões de bucketing do Amazon EMR EMR 6.x e 5.x Hive

EMR 5.x usa o OOS Apache Hive 2, enquanto no EMR 6.x usa OOS o Apache Hive 3. O Hive2 de código aberto usa o Bucketing versão 1, enquanto o Hive3 de código aberto usa o Bucketing versão 2. Essa diferença de versão de compartimentação entre o Hive 2 (EMR5.x) e o Hive 3 (EMR6.x) significa que o hashing de compartimentação do Hive funciona de forma diferente. Veja o exemplo abaixo.

A tabela a seguir é um exemplo criado em EMR 6.x e EMR 5.x, respectivamente.

```
-- 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/';
```

Inserindo os mesmos dados em EMR 6.x e 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');
```

Verificar a localização do S3 mostra que o nome do arquivo de compartimento é diferente, porque a função de hash é diferente entre EMR 6.x (Hive 3) e 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
```

Você também pode ver a diferença de versão executando o seguinte comando no Hive CLI na versão EMR 6.x. Observe que ele retorna a versão 2 do bucketing.

```
hive> DESCRIBE FORMATTED test_bucketing;
...
Table Parameters:
  bucketing_version      2
...
```

- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSH como usuário do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do hadoop.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo keytab está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

- Quando você usa o Spark com a formatação de localização de partições do Hive para ler dados no Amazon S3 e executa o Spark nas EMR versões 5.30.0 a 5.36.0 e 6.2.0 a 6.9.0 da Amazon, você pode encontrar um problema que impede seu cluster de ler os dados corretamente. Isso poderá acontecer se suas partições tiverem todas as características a seguir:
 - Duas ou mais partições são verificadas na mesma tabela.

- Pelo menos um caminho de diretório de partição é um prefixo de pelo menos outro caminho de diretório de partição, por exemplo, `s3://bucket/table/p=a` é um prefixo de `s3://bucket/table/p=a b`.
- O primeiro caractere que segue o prefixo no outro diretório de partição tem um valor UTF -8 menor que o / caractere (U+002F). Por exemplo, o caractere de espaço (U+0020) que ocorre entre a e b em `s3://bucket/table/p=a b` se enquadra nessa categoria. Observe que existem 14 outros caracteres que não são de controle: `!"#$%&'()*+,-.` Para obter mais informações, consulte a [tabela de codificação UTF -8 e caracteres Unicode](#).

Como solução alternativa para esse problema, defina a configuração `spark.sql.sources.fastS3PartitionDiscovery.enabled` como `false` na classificação `spark-defaults`.

Versões de componente 6.2.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.16.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
emr-goodies	3.1.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.5.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-notebook-env	1.0.0	Ambiente Conda para bloco de anotações do emr, que inclui o jupyter enterprise gateway
emr-s3-dist-cp	2.16.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	2.0.0	EMRConector S3Select
emrfs	2.44.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.11.2	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
flink-jobmanager-config	1.11.2	Gerenciando recursos em EMR nós para o Apache JobManager Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.

Componente	Version (Versão)	Descrição
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	3.2.1-amzn-2	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	3.2.1-amzn-2	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	3.2.1-amzn-2	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	3.2.1-amzn-2	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	3.2.1-amzn-2	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httpfs-server	3.2.1-amzn-2	HTTPendpoint para HDFS operações.
hadoop-kms-server	3.2.1-amzn-2	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API

Componente	Version (Versão)	Descrição
hadoop-mapred	3.2.1-amzn-2	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	3.2.1-amzn-2	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	3.2.1-amzn-2	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	3.2.1-amzn-2	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	2.2.6-amzn-0	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	2.2.6-amzn-0	Serviço para atender a uma ou mais HBase regiões.
hbase-client	2.2.6-amzn-0	HBasecliente de linha de comando.
hbase-rest-server	2.2.6-amzn-0	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	2.2.6-amzn-0	Serviço que fornece um endpoint Thrift para. HBase

Componente	Version (Versão)	Descrição
hcatalog-client	3.1.2-amzn-3	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	3.1.2-amzn-3	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	3.1.2-amzn-3	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	3.1.2-amzn-3	O cliente da linha de comando do Hive.
hive-hbase	3.1.2-amzn-3	Cliente do Hive-hbase.
hive-metastore-server	3.1.2-amzn-3	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	3.1.2-amzn-3	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.6.0-amzn-1	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.
hudi-presto	0.6.0-amzn-1	Biblioteca de pacotes para executar o Presto com o Hudi.

Componente	Version (Versão)	Descrição
hudi-prestosql	0.6.0-amzn-1	Biblioteca de pacotes para executar o Presto SQL com o Hudi.
hudi-spark	0.6.0-amzn-1	Biblioteca de pacotes para executar o Spark com o Hudi.
hue-server	4.8.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	1.1.0	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.7.0-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mxnet	1.7.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mariadb-server	5.5.64+	Servidor de banco de dados MariaDB.
nvidia-cuda	10.1.243	Drivers NVIDIA e toolkit CUDA
oozie-client	5.2.0	O cliente da linha de comando do Oozie.
oozie-server	5.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.

Componente	Version (Versão)	Descrição
opencv	4.4.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	5.0.0- HBase -2,0	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	5.0.0- HBase -2,0	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.238.3-amzn-1	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.238.3-amzn-1	O serviço que executa partes de uma consulta.
presto-client	0.238.3-amzn-1	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
prestosql-coordinator	343	O serviço que aceita consultas e gerencia a execução de consultas entre os prestosql-workers.
prestosql-worker	343	O serviço que executa partes de uma consulta.

Componente	Version (Versão)	Descrição
prestosql-client	343	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.3	O projeto R para computação estatística
ranger-kms-server	2.0.0	Sistema de gerenciamento de chaves Apache Ranger
spark-client	3.0.1-amzn-0	Os clientes da linha de comando do Spark.
spark-history-server	3.0.1-amzn-0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	3.0.1-amzn-0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	3.0.1-amzn-0	Bibliotecas do Apache Spark necessárias para escravos. YARN
spark-rapids	0.2.0	RAPIDSPlugin Nvidia Spark que acelera o Apache Spark com o. GPUs

Componente	Version (Versão)	Descrição
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	2.3.1	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.2	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.41+	HTTPServidor Apache.
zeppelin-server	0.9.0-preview1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.14	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.14	ZooKeeper cliente de linha de comando.

Classificações de configuração 6.2.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Ocorrem ações de reconfiguração quando você especifica uma configuração para grupos de instâncias em um cluster em execução. A Amazon EMR só inicia ações de reconfiguração para as

classificações que você modifica. Para ter mais informações, consulte [Reconfigurar um grupo de instâncias de um cluster em execução](#).

Classificações do emr-6.2.0

Classificações	Descrição	Ações de reconfiguração
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.	Restarts the Resource Manager service.
container-executor	Altere os valores no arquivo container-executor.cfg YARN do Hadoop.	Not available.
container-log4j	Altere os valores no arquivo container-log4j.properties YARN do Hadoop.	Not available.
core-site	Alterar os valores no arquivo core-site.xml do 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	Alterar as configurações relacionadas ao docker.	Not available.
emrfs-site	Altere EMRFS as configurações.	Restarts the Hadoop HDFS services Namenode,

Classificações	Descrição	Ações de reconfiguração
		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	Alterar as configurações do <code>flink-conf.yaml</code> .	Not available.
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.	Not available.
flink-log4j-yarn-session	Altere as configurações do <code>Flink log4j-yarn-session.properties</code> .	Not available.
flink-log4j-cli	Alterar as configurações de <code>log4j-cli.properties</code> no Flink.	Not available.

Classificações	Descrição	Ações de reconfiguração
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do 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	Alterar os valores no arquivo log4j.properties do 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	Alterar a configuração do servidor SSL no Hadoop	Not available.
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop	Not available.

Classificações	Descrição	Ações de reconfiguração
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Mude os valores no HBase ambiente.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.	Not available.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.	This classification should not be reconfigured.
hdfs-env	Mude os valores no HDFS ambiente.	Restarts Hadoop HDFS ZKFC.

Classificações	Descrição	Ações de reconfiguração
hdfs-site	Altere os valores no HDFS hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Mude os valores no HCatalog ambiente.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat server.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat	Restarts Hive WebHCat server.
hive	Configurações EMR selecionadas pela Amazon para o Apache Hive.	Sets configurations to launch Hive LLAP service.
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.	Not available.

Classificações	Descrição	Ações de reconfiguração
hive-parquet-logging	Alterar os valores no arquivo <code>parquet-logging.properties</code> do Hive.	Not available.
hive-env	Alterar os valores no ambiente do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive <code>schemaTool CLI</code> commands to verify <code>hive-metastore</code> .
hive-exec-log4j2	Altere os valores no arquivo <code>hive-exec-log 4j2.properties</code> do Hive.	Not available.
hive-llap-daemon-log4j2	Altere os valores no arquivo <code>llap-daemon-log 4j2.properties</code> do Hive.	Not available.
hive-log4j2	Alterar os valores no arquivo <code>hive-log4j2.properties</code> do Hive.	Not available.
hive-site	Alterar os valores no arquivo <code>hive-site.xml</code> do Hive.	Restarts HiveServer2, HiveMetastore, and Hive HCatalog-Server. Runs Hive <code>schemaTool CLI</code> commands to verify <code>hive-metastore</code> . Also restarts Oozie and Zeppelin.
hiveserver2-site	Alterar os valores no arquivo <code>hiveserver2-site.xml</code> do Hive Server2.	Not available.
hue-ini	Alterar os valores no arquivo <code>ini</code> do Hue	Restarts Hue. Also activates Hue <code>config override CLI</code> commands to pick up new configurations.

Classificações	Descrição	Ações de reconfiguração
httpfs-env	Mude os valores no HTTPFS ambiente.	Restarts Hadoop Httpfs service.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.	Not available.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.	Not available.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Altere os valores no ambiente do Hudi.	Not available.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.	Not available.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.	Not available.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.	Not available.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.	Not available.

Classificações	Descrição	Ações de reconfiguração
livy-conf	Alterar os valores no arquivo livy.conf do Livy.	Restarts Livy Server.
livy-env	Alterar os valores no ambiente do Livy.	Restarts Livy Server.
livy-log4j	Alterar as configurações de log4j.properties no Livy.	Restarts Livy Server.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Alterar os valores no ambiente do Oozie.	Restarts Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.	Restarts Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.	Not available.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.	Not available.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.	Restarts Phoenix-QueryServer.

Classificações	Descrição	Ações de reconfiguração
phoenix-metrics	Alterar os valores no arquivo <code>hadoop-metrics2-phoenix.properties</code> do Phoenix.	Not available.
pig-env	Alterar os valores no ambiente do Pig.	Not available.
pig-properties	Alterar os valores no arquivo <code>pig.properties</code> do Pig.	Restarts Oozie.
pig-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Pig.	Not available.
presto-log	Alterar os valores no arquivo <code>log.properties</code> do Presto.	Restarts Presto-Server (for PrestoDB)
presto-config	Alterar os valores no arquivo <code>config.properties</code> do Presto.	Restarts Presto-Server (for PrestoDB)
presto-password-authenticator	Altere os valores no arquivo <code>password-authenticator.properties</code> do Presto.	Not available.
presto-env	Altere os valores no arquivo <code>presto-env.sh</code> do Presto.	Restarts Presto-Server (for PrestoDB)
presto-node	Altere os valores no arquivo <code>node.properties</code> do Presto.	Not available.
presto-connector-blackhole	Alterar os valores no arquivo <code>blackhole.properties</code> do Presto.	Not available.
presto-connector-cassandra	Alterar os valores no arquivo <code>cassandra.properties</code> do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.	Restarts Presto-Server (for PrestoDB)
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.	Not available.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.	Not available.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.	Not available.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.	Not available.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.	Not available.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.	Not available.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.	Not available.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.	Not available.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.	Not available.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.	Not available.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.	Not available.
prestosql-log	Alterar os valores no arquivo log.properties do Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-config	Alterar os valores no arquivo config.properties do Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-env	Altere os valores no arquivo presto-env.sh do Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-node	Altere os valores no arquivo SQL node.properties do Presto.	Not available.
prestosql-connector-blackhole	Altere os valores no arquivo SQL blackhole.properties do Presto.	Not available.
prestosql-connector-cassandra	Altere os valores no arquivo SQL cassandra.properties do Presto.	Not available.
prestosql-connector-hive	Altere os valores no arquivo SQL hive.properties do Presto.	Restarts Presto-Server (for PrestoSQL)
prestosql-connector-jmx	Altere os valores no arquivo SQL jmx.properties do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
prestosql-connector-kafka	Altere os valores no arquivo SQL kafka.properties do Presto.	Not available.
prestosql-connector-localfile	Altere os valores no arquivo SQL localfile.properties do Presto.	Not available.
prestosql-connector-memory	Altere os valores no arquivo SQL memory.properties do Presto.	Not available.
prestosql-connector-mongodb	Altere os valores no arquivo SQL mongodb.properties do Presto.	Not available.
prestosql-connector-mysql	Altere os valores no arquivo SQL mysql.properties do Presto.	Not available.
prestosql-connector-postgresql	Altere os valores no arquivo SQL postgresql.properties do Presto.	Not available.
prestosql-connector-raptor	Altere os valores no arquivo SQL raptor.properties do Presto.	Not available.
prestosql-connector-redis	Altere os valores no arquivo SQL redis.properties do Presto.	Not available.
prestosql-connector-redshift	Altere os valores no arquivo SQL redshift.properties do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
prestosql-connector-tpch	Altere os valores no arquivo SQL tpch.properties do Presto.	Not available.
prestosql-connector-tpcds	Altere os valores no arquivo SQL tpcds.properties do Presto.	Not available.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Altere os valores no arquivo kms-log4j.properties do Ranger. KMS	Not available.
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.	Not available.
spark	Configurações EMR seleciona das pela Amazon para o Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Alterar os valores no ambiente do Spark.	Restarts Spark history server and Spark thrift server.

Classificações	Descrição	Ações de reconfiguração
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.	Not available.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Alterar os valores no ambiente do Sqoop.	Not available.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.	Not available.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.	Not available.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.	Restart Oozie.
yarn-env	Mude os valores no YARN ambiente.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.

Classificações	Descrição	Ações de reconfiguração
zeppelin-env	Alterar os valores no ambiente do Zeppelin.	Restarts Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.	Restarts Zookeeper server.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.	Restarts Zookeeper server.

Amazon EMR versão 6.1.1

Versões 6.1.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [PrestoSQL](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
AWS SDK para Java	1.11.828	1.11.828	1.11.711	1.11.711

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12.10	2.12.10	2.12.10	2.11.12
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
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

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
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
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

Notas da versão 6.1.1

Esta é uma versão para corrigir problemas com o Amazon EMR Scaling quando ele falha na escalabilidade ou redução de escala de um cluster com sucesso ou causa falhas no aplicativo.

Alterações, melhorias e problemas resolvidos

- Corrigido um problema em que as solicitações de escalabilidade falhavam em um cluster grande e altamente utilizado quando os daemons no cluster da EMR Amazon estavam executando atividades de verificação de integridade, como coletar o estado do nó e o estado do YARN nó. HDFS Isso estava acontecendo porque os daemons no cluster não conseguiam comunicar os dados do status de saúde de um nó aos componentes internos da Amazon. EMR
- Daemons EMR no cluster aprimorados para rastrear corretamente os estados dos nós quando os endereços IP são reutilizados para melhorar a confiabilidade durante as operações de escalabilidade.

- [SPARK-29683](#). Corrigido um problema em que ocorriam falhas de trabalho durante a redução da escala verticalmente do cluster, pois o Spark presumia que todos os nós disponíveis estavam na lista de negação.
- [YARN-9011](#). Corrigido um problema em que falhas de trabalho ocorriam devido a uma condição de corrida no YARN descomissionamento quando o cluster tentava aumentar ou diminuir a escala.
- Foi corrigido o problema com falhas de etapas ou tarefas durante a escalabilidade do cluster, garantindo que os estados dos nós fossem sempre consistentes entre os EMR daemons no cluster da Amazon e/. YARN HDFS
- Corrigido um problema em que operações de cluster, como redução de escala e envio de etapas, falhavam em EMR clusters da Amazon habilitados com a autenticação Kerberos. Isso ocorreu porque o daemon no EMR cluster da Amazon não renovou o tíquete Kerberos, que é necessário para se comunicar de forma segura com/em execução no nó primário. HDFS YARN
- As EMR versões mais recentes da Amazon corrigem o problema com um limite menor de “Máximo de arquivos abertos” em relação aos mais antigos AL2 na AmazonEMR. As EMR versões 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 e posteriores agora incluem uma correção permanente com uma configuração mais alta de “Máximo de arquivos abertos”.
- HTTPS agora está habilitado por padrão para os repositórios Amazon Linux. Se você estiver usando uma VPCE política do Amazon S3 para restringir o acesso a buckets específicos, deverá adicionar o novo bucket Amazon Linux ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` à sua política (`$region` substitua pela região onde está o endpoint). Para obter mais informações, consulte esse tópico nos fóruns de AWS discussão. [Anúncio: o Amazon Linux 2 agora oferece suporte à capacidade de uso HTTPS ao se conectar a repositórios de pacotes.](#)

Versões de componente 6.1.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto

nomeado myapp-component com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como 2.2-amzn-2.

Componente	Version (Versão)	Descrição
aws-sagemaker-spark-sdk	1.3.0	Amazon SageMaker Spark SDK
emr-ddb	4.14.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	3.1.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.5.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.14.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	2.0.0	EMRConector S3Select
emrfs	2.42.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.11.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.

Componente	Version (Versão)	Descrição
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	3.2.1-amzn-1.1	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	3.2.1-amzn-1.1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	3.2.1-amzn-1.1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	3.2.1-amzn-1.1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	3.2.1-amzn-1.1	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httpfs-server	3.2.1-amzn-1.1	HTTPendpoint para HDFS operações.
hadoop-kms-server	3.2.1-amzn-1.1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API

Componente	Version (Versão)	Descrição
hadoop-mapred	3.2.1-amzn-1.1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	3.2.1-amzn-1.1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	3.2.1-amzn-1.1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	3.2.1-amzn-1.1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	2.2.5	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	2.2.5	Serviço para atender a uma ou mais HBase regiões.
hbase-client	2.2.5	HBasecliente de linha de comando.
hbase-rest-server	2.2.5	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	2.2.5	Serviço que fornece um endpoint Thrift para. HBase

Componente	Version (Versão)	Descrição
hcatalog-client	3.1.2-amzn-2	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	3.1.2-amzn-2	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	3.1.2-amzn-2	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	3.1.2-amzn-2	O cliente da linha de comando do Hive.
hive-hbase	3.1.2-amzn-2	Cliente do Hive-hbase.
hive-metastore-server	3.1.2-amzn-2	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	3.1.2-amzn-2	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.5.2-incubating-amzn-2	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.
hudi-presto	0.5.2-incubating-amzn-2	Biblioteca de pacotes para executar o Presto com o Hudi.

Componente	Version (Versão)	Descrição
hudi-prestosql	0.5.2-incubating-amzn-2	Biblioteca de pacotes para executar o Presto SQL com o Hudi.
hudi-spark	0.5.2-incubating-amzn-2	Biblioteca de pacotes para executar o Spark com o Hudi.
hue-server	4.7.1	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	1.1.0	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.7.0-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mxnet	1.6.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mariadb-server	5.5.64+	Servidor de banco de dados MariaDB.
nvidia-cuda	9.2.88	Drivers NVIDIA e toolkit CUDA
oozie-client	5.2.0	O cliente da linha de comando do Oozie.
oozie-server	5.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.

Componente	Version (Versão)	Descrição
opencv	4.3.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	5.0.0- HBase -2,0	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	5.0.0- HBase -2,0	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.232	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.232	O serviço que executa partes de uma consulta.
presto-client	0.232	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
prestosql-coordinator	338	O serviço que aceita consultas e gerencia a execução de consultas entre os prestosql-workers.
prestosql-worker	338	O serviço que executa partes de uma consulta.

Componente	Version (Versão)	Descrição
prestosql-client	338	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.3	O projeto R para computação estatística
ranger-kms-server	2.0.0	Sistema de gerenciamento de chaves Apache Ranger
spark-client	3.0.0-amzn-0.1	Os clientes da linha de comando do Spark.
spark-history-server	3.0.0-amzn-0.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	3.0.0-amzn-0.1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	3.0.0-amzn-0.1	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.

Componente	Version (Versão)	Descrição
tensorflow	2.1.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.2	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.41+	HTTPServidor Apache.
zeppelin-server	0.9.0-preview1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.14	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.14	ZooKeeper cliente de linha de comando.

Classificações de configuração 6.1.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-6.1.1

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.

Classificações	Descrição
container-executor	Altere os valores no arquivo container-executor .cfg YARN do Hadoop.
container-log4j	Altere os valores no arquivo container-log4j.properties YARN do Hadoop.
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4j-yarn-session.properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.

Classificações	Descrição
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-env	Mude os valores no HDFS ambiente.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive	Configurações EMR selecionadas pela Amazon para o Apache Hive.

Classificações	Descrição
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.

Classificações	Descrição
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
hudi-env	Altere os valores no ambiente do Hudi.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.

Classificações	Descrição
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.

Classificações	Descrição
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.

Classificações	Descrição
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.
prestosql-log	Alterar os valores no arquivo log.properties do Presto.
prestosql-config	Alterar os valores no arquivo config.properties do Presto.
prestosql-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
prestosql-env	Altere os valores no arquivo presto-env.sh do Presto.
prestosql-node	Altere os valores no arquivo SQL node.properties do Presto.
prestosql-connector-blackhole	Altere os valores no arquivo SQL blackhole.properties do Presto.
prestosql-connector-cassandra	Altere os valores no arquivo SQL cassandra.properties do Presto.
prestosql-connector-hive	Altere os valores no arquivo SQL hive.properties do Presto.
prestosql-connector-jmx	Altere os valores no arquivo SQL jmx.properties do Presto.
prestosql-connector-kafka	Altere os valores no arquivo SQL kafka.properties do Presto.
prestosql-connector-localfile	Altere os valores no arquivo SQL localfile.properties do Presto.
prestosql-connector-memory	Altere os valores no arquivo SQL memory.properties do Presto.

Classificações	Descrição
prestosql-connector-mongodb	Altere os valores no arquivo SQL mongodb.properties do Presto.
prestosql-connector-mysql	Altere os valores no arquivo SQL mysql.properties do Presto.
prestosql-connector-postgresql	Altere os valores no arquivo SQL postgresql.properties do Presto.
prestosql-connector-raptor	Altere os valores no arquivo SQL raptor.properties do Presto.
prestosql-connector-redis	Altere os valores no arquivo SQL redis.properties do Presto.
prestosql-connector-redshift	Altere os valores no arquivo SQL redshift.properties do Presto.
prestosql-connector-tpch	Altere os valores no arquivo SQL tpch.properties do Presto.
prestosql-connector-tpcds	Altere os valores no arquivo SQL tpcds.properties do Presto.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.
ranger-kms-log4j	Altere os valores no arquivo kms-log4j.properties do Ranger. KMS
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.

Classificações	Descrição
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.

Classificações	Descrição
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 6.1.0

Versões 6.1.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [PrestoSQL](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
AWS SDK para Java	1.11.828	1.11.828	1.11.711	1.11.711
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12.10	2.12.10	2.12.10	2.11.12
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

Notas da versão 6.1.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 6.1.0 da Amazon. As alterações são referentes à versão 6.0.0.

Data da versão inicial: 4 de setembro de 2020

Data da última atualização: 15 de outubro de 2020

Aplicações compatíveis

- AWS SDK for Java versão 1.11.828
- Flink versão 1.11.0
- Ganglia versão 3.7.2
- Hadoop versão 3.2.1-amzn-1
- HBase versão 2.2.5
- HBase-operator-tools 1.0.0
- HCatalog versão 3.1.2-amzn-0
- Hive versão 3.1.2-amzn-1
- Hudi versão 0.5.2-incubating
- Hue versão 4.7.1

- JupyterHub versão 1.1.0
- Livy versão 0.7.0
- MXNet versão 1.6.0
- Oozie versão 5.2.0
- Phoenix versão 5.0.0
- Presto versão 0.232
- Presto SQL versão 3.38
- Spark versão 3.0.0-amzn-0
- TensorFlow versão 2.1.0
- Zeppelin versão 0.9.0-preview1
- Zookeeper versão 3.4.14
- Conectores e drivers: DynamoDB Connector 4.14.0

Novos atributos

- ARMOs tipos de instância são suportados a partir da Amazon EMR versão 5.30.0 e da Amazon EMR versão 6.1.0.
- Os tipos de instância de uso geral M6g são suportados a partir das EMR versões 6.1.0 e 5.30.0 da Amazon. Para obter mais informações, consulte [Tipos de instância compatíveis](#) no Amazon EMR Management Guide.
- O recurso de grupos de EC2 posicionamento é suportado a partir da EMR versão 5.23.0 da Amazon como uma opção para vários clusters de nós primários. Atualmente, somente os tipos de nós primários são compatíveis com o atributo grupo de posicionamento e a estratégia SPREAD é aplicada a estes nós primários. A estratégia SPREAD posiciona um pequeno grupo de instâncias em um hardware subjacente separado para evitar a perda de múltiplos nós primários em caso de falha de hardware. Para obter mais informações, consulte [EMRIntegração com o EC2 Placement Group](#) no Amazon EMR Management Guide.
- Escalabilidade gerenciada — Com a EMR versão 6.1.0 da Amazon, você pode permitir que a escalabilidade EMR gerenciada da Amazon aumente ou diminua automaticamente o número de instâncias ou unidades em seu cluster com base na carga de trabalho. A Amazon avalia EMR continuamente as métricas do cluster para tomar decisões de escalabilidade que otimizem seus clusters em termos de custo e velocidade. O Managed Scaling também está disponível na Amazon EMR versão 5.30.0 e posterior, exceto 6.0.0. Para obter mais informações, consulte [Scaling Cluster Resources](#) no Amazon EMR Management Guide.

- A SQL versão 338 do Presto é compatível com EMR 6.1.0. Para obter mais informações, consulte [Presto](#).
 - O Presto SQL é suportado somente na EMR versão 6.1.0 e versões posteriores, não na EMR 6.0.0 ou 5.x. EMR
 - O nome da aplicação, `Presto`, continua a ser usado para instalar o PrestoDB em clusters. Para instalar o Presto SQL em clusters, use o nome `PrestoSQL` do aplicativo.
 - Você pode instalar o PrestoDB ou o SQL Presto, mas não pode instalar os dois em um único cluster. Se o PrestoDB e o SQL Presto forem especificados ao tentar criar um cluster, ocorrerá um erro de validação e a solicitação de criação do cluster falhará.
 - O Presto SQL é compatível com clusters single-master e multimaster. Em clusters com vários mestres, é necessário um metastore externo do Hive para executar o Presto ou o PrestoDB. SQL Consulte [Aplicativos compatíveis em um EMR cluster com vários nós primários](#).
- ECRsuporte à autenticação automática no Apache Hadoop e no Apache Spark com Docker: os usuários do Spark podem usar imagens do Docker Hub e do Amazon Elastic Container Registry (AmazonECR) para definir dependências de ambiente e biblioteca.

[Configure o Docker](#) e [execute aplicativos Spark com o Docker usando o Amazon 6.x](#). EMR

- EMRsuporta ACID transações do Apache Hive: o Amazon EMR 6.1.0 adiciona suporte às transações do Hive para que esteja em conformidade com ACID as propriedades de um banco de dados. ACID Com esse atributo, você pode executar as operações `INSERT`, `UPDATE`, `DELETE`, e `MERGE` em tabelas gerenciadas do Hive com dados no Amazon Simple Storage Service (Amazon S3). Esse é um recurso fundamental para casos de uso como ingestão de streaming, atualização de dados, uso de atualizações em massa e mudanças lentas de dimensões. `MERGE` Para obter mais informações, incluindo exemplos de configuração e casos de uso, consulte [Amazon EMR suporta transações do Apache Hive ACID](#).

Alterações, melhorias e problemas resolvidos

- Esta é uma versão para corrigir problemas com o Amazon EMR Scaling quando ele falha na escalabilidade ou redução de escala de um cluster com sucesso ou causa falhas no aplicativo.
- Corrigido um problema em que as solicitações de escalabilidade falhavam em um cluster grande e altamente utilizado quando os daemons no cluster da EMR Amazon estavam executando atividades de verificação de integridade, como coletar o estado do nó e o estado do YARN nó. HDFS Isso estava acontecendo porque os daemons no cluster não conseguiam comunicar os dados do status de saúde de um nó aos componentes internos da Amazon. EMR

- Daemons EMR no cluster aprimorados para rastrear corretamente os estados dos nós quando os endereços IP são reutilizados para melhorar a confiabilidade durante as operações de escalabilidade.
- [SPARK-29683](#). Corrigido um problema em que ocorriam falhas de trabalho durante a redução da escala verticalmente do cluster, pois o Spark presumia que todos os nós disponíveis estavam na lista de negação.
- [YARN-9011](#). Corrigido um problema em que falhas de trabalho ocorriam devido a uma condição de corrida no YARN descomissionamento quando o cluster tentava aumentar ou diminuir a escala.
- Foi corrigido o problema com falhas de etapas ou tarefas durante a escalabilidade do cluster, garantindo que os estados dos nós fossem sempre consistentes entre os EMR daemons no cluster da Amazon e/. YARN HDFS
- Corrigido um problema em que operações de cluster, como redução de escala e envio de etapas, falhavam em EMR clusters da Amazon habilitados com a autenticação Kerberos. Isso ocorreu porque o daemon no EMR cluster da Amazon não renovou o tíquete Kerberos, que é necessário para se comunicar de forma segura com/em execução no nó primário. HDFS YARN
- As EMR versões mais recentes da Amazon corrigem o problema com um limite menor de “Máximo de arquivos abertos” em relação aos mais antigos AL2 na AmazonEMR. As EMR versões 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 e posteriores agora incluem uma correção permanente com uma configuração mais alta de “Máximo de arquivos abertos”.
- O Apache Flink não é suportado no EMR 6.0.0, mas é suportado no EMR 6.1.0 com o Flink 1.11.0. Esta é a primeira versão do Flink a oficialmente oferecer suporte ao Hadoop 3. Consulte o [Anúncio de versão do Apache Flink 1.11.0](#).
- O Ganglia foi removido dos pacotes padrão EMR 6.1.0.

Problemas conhecidos

- Limite inferior de “Máximo de arquivos abertos” em versões mais antigas AL2 [corrigido em versões mais recentes]. EMRLançamentos da Amazon: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 e emr-6.2.0 são baseados em versões mais antigas do Linux ofAmazon 2 (AL2, que têm uma configuração de limite inferior para “Máximo de arquivos abertos” quando os clusters da Amazon são criados com o padrão. EMR AMI As EMR versões 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 e posteriores incluem uma correção permanente com uma configuração maior de “Máximo de arquivos abertos”. Versões com o limite inferior de arquivos abertos causam o erro “Muitos arquivos abertos” ao ser enviado um trabalho do Spark. Nas versões afetadas, o padrão da Amazon AMI tem uma configuração EMR padrão de ulimit

de 4096 para “Máximo de arquivos abertos”, que é inferior ao limite de 65536 arquivos no Linux 2. latestAmazon AMI A configuração inferior de ulimit para “Máximo de arquivos abertos” causa falhas em trabalhos do Spark quando o driver e o executor do Spark tentam abrir mais de 4096 arquivos. Para corrigir o problema, a Amazon EMR tem um script de ação de bootstrap (BA) que ajusta a configuração ulimit na criação do cluster.

Se você estiver usando uma EMR versão mais antiga da Amazon que não tem a solução permanente para esse problema, a solução alternativa a seguir permite que você defina explicitamente o ulimit do controlador de instância para um máximo de 65536 arquivos.

Defina explicitamente um ulimit na linha de comando

1. Edite `/etc/systemd/system/instance-controller.service` para adicionar os seguintes parâmetros à seção Serviço.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Reiniciar InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Defina um ulimit usando a ação de bootstrap (BA)

Você também pode usar um script de ação de bootstrap (BA) para configurar o ulimit instance-controller para 65536 arquivos na criação do 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**

O Amazon EMR 6.1.0 e 6.2.0 incluem um problema de desempenho que pode afetar criticamente todas as operações de inserção, atualização e exclusão do Hudi. Se você planeja usar o Hudi com o Amazon EMR 6.1.0 ou 6.2.0, entre em contato com o AWS suporte para obter um Hudi corrigido. RPM

- Se você definir uma configuração personalizada de coleta de lixo com `spark.driver.extraJavaOptions` e `spark.executor.extraJavaOptions`, isso resultará na falha de inicialização do driver/executor com a versão EMR 6.1 devido à configuração conflitante da coleta de lixo. Com a EMR versão 6.1.0, você deve especificar uma configuração personalizada de coleta de lixo do Spark para drivers e executores com as propriedades e em vez disso `spark.driver.defaultJavaOptions` e `spark.executor.defaultJavaOptions`. Leia mais em [Apache Spark Runtime Environment](#) e [Configurando a coleta de lixo do Spark](#) na Amazon EMR 6.1.0.
- Usar o Pig com o Oozie (e dentro do Hue, já que o Hue usa ações do Oozie para executar scripts do Pig) gera um erro em que uma biblioteca nativa Izo não pode ser carregada. Essa mensagem de erro é informativa e não impede a execução do Pig.
- Suporte de simultaneidade do Hudi: atualmente, o Hudi não é compatível com gravações simultâneas em uma única tabela do Hudi. Além disso, o Hudi reverte todas as alterações feitas por gravadores em andamento antes de permitir que um novo gravador seja iniciado. As gravações simultâneas podem interferir nesse mecanismo e introduzir condições de corrida, o que pode causar corrupção de dados. Você deve garantir que, como parte do seu fluxo de trabalho de processamento de dados, só exista um gravador do Hudi operando em uma tabela do Hudi em qualquer instante. O Hudi permite vários leitores simultâneos operando na mesma tabela do Hudi.
- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema

de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSH como usuário do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do hadoop.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo keytab está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

- Há um problema no Amazon EMR 6.1.0 que afeta os clusters que executam o Presto. Depois de um longo período (dias), o cluster pode gerar erros, como “su: failed to execute /bin/bash: Resource temporarily unavailable” ou “shell request failed on channel 0”. Esse problema é causado por um EMR processo interno da Amazon (InstanceController) que está gerando muitos processos leves (LWP), o que acaba fazendo com que o usuário do Hadoop exceda o limite de `nproc`. Isso impede que o usuário abra processos adicionais. A solução para esse problema é atualizar para EMR 6.2.0.

Versões de componente 6.1.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado myapp-component com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como 2.2-amzn-2.

Componente	Version (Versão)	Descrição
aws-sagemaker-spark-sdk	1.3.0	Amazon SageMaker Spark SDK
emr-ddb	4.14.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	3.1.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.5.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.14.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	2.0.0	EMRConector S3Select
emrfs	2.42.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.11.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop,

Componente	Version (Versão)	Descrição
		juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	3.2.1-amzn-1	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	3.2.1-amzn-1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	3.2.1-amzn-1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	3.2.1-amzn-1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	3.2.1-amzn-1	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httpfs-server	3.2.1-amzn-1	HTTPendpoint para HDFS operações.
hadoop-kms-server	3.2.1-amzn-1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API

Componente	Version (Versão)	Descrição
hadoop-mapred	3.2.1-amzn-1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	3.2.1-amzn-1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	3.2.1-amzn-1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	3.2.1-amzn-1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	2.2.5	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	2.2.5	Serviço para atender a uma ou mais HBase regiões.
hbase-client	2.2.5	HBasecliente de linha de comando.
hbase-rest-server	2.2.5	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	2.2.5	Serviço que fornece um endpoint Thrift para. HBase

Componente	Version (Versão)	Descrição
hcatalog-client	3.1.2-amzn-2	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	3.1.2-amzn-2	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	3.1.2-amzn-2	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	3.1.2-amzn-2	O cliente da linha de comando do Hive.
hive-hbase	3.1.2-amzn-2	Cliente do Hive-hbase.
hive-metastore-server	3.1.2-amzn-2	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	3.1.2-amzn-2	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.5.2-incubating-amzn-2	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.
hudi-presto	0.5.2-incubating-amzn-2	Biblioteca de pacotes para executar o Presto com o Hudi.

Componente	Version (Versão)	Descrição
hudi-prestosql	0.5.2-incubating-amzn-2	Biblioteca de pacotes para executar o Presto SQL com o Hudi.
hudi-spark	0.5.2-incubating-amzn-2	Biblioteca de pacotes para executar o Spark com o Hudi.
hue-server	4.7.1	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	1.1.0	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.7.0-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mxnet	1.6.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mariadb-server	5.5.64+	Servidor de banco de dados MariaDB.
nvidia-cuda	9.2.88	Drivers NVIDIA e toolkit CUDA
oozie-client	5.2.0	O cliente da linha de comando do Oozie.
oozie-server	5.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.

Componente	Version (Versão)	Descrição
opencv	4.3.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	5.0.0- HBase -2,0	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	5.0.0- HBase -2,0	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.232	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.232	O serviço que executa partes de uma consulta.
presto-client	0.232	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
prestosql-coordinator	338	O serviço que aceita consultas e gerencia a execução de consultas entre os prestosql-workers.
prestosql-worker	338	O serviço que executa partes de uma consulta.

Componente	Version (Versão)	Descrição
prestosql-client	338	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.3	O projeto R para computação estatística
ranger-kms-server	2.0.0	Sistema de gerenciamento de chaves Apache Ranger
spark-client	3.0.0-amzn-0	Os clientes da linha de comando do Spark.
spark-history-server	3.0.0-amzn-0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	3.0.0-amzn-0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	3.0.0-amzn-0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.

Componente	Version (Versão)	Descrição
tensorflow	2.1.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.2	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.41+	HTTPServidor Apache.
zeppelin-server	0.9.0-preview1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.14	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.14	ZooKeeper cliente de linha de comando.

Classificações de configuração 6.1.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-6.1.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.

Classificações	Descrição
container-executor	Altere os valores no arquivo container-executor .cfg YARN do Hadoop.
container-log4j	Altere os valores no arquivo container-log4j.properties YARN do Hadoop.
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4j-yarn-session.properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.

Classificações	Descrição
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-env	Mude os valores no HDFS ambiente.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive	Configurações EMR selecionadas pela Amazon para o Apache Hive.

Classificações	Descrição
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.

Classificações	Descrição
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
hudi-env	Altere os valores no ambiente do Hudi.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.

Classificações	Descrição
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.

Classificações	Descrição
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.

Classificações	Descrição
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.
prestosql-log	Alterar os valores no arquivo log.properties do Presto.
prestosql-config	Alterar os valores no arquivo config.properties do Presto.
prestosql-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
prestosql-env	Altere os valores no arquivo presto-env.sh do Presto.
prestosql-node	Altere os valores no arquivo SQL node.properties do Presto.
prestosql-connector-blackhole	Altere os valores no arquivo SQL blackhole.properties do Presto.
prestosql-connector-cassandra	Altere os valores no arquivo SQL cassandra.properties do Presto.
prestosql-connector-hive	Altere os valores no arquivo SQL hive.properties do Presto.
prestosql-connector-jmx	Altere os valores no arquivo SQL jmx.properties do Presto.
prestosql-connector-kafka	Altere os valores no arquivo SQL kafka.properties do Presto.
prestosql-connector-localfile	Altere os valores no arquivo SQL localfile.properties do Presto.
prestosql-connector-memory	Altere os valores no arquivo SQL memory.properties do Presto.

Classificações	Descrição
prestosql-connector-mongodb	Altere os valores no arquivo SQL mongodb.properties do Presto.
prestosql-connector-mysql	Altere os valores no arquivo SQL mysql.properties do Presto.
prestosql-connector-postgresql	Altere os valores no arquivo SQL postgresql.properties do Presto.
prestosql-connector-raptor	Altere os valores no arquivo SQL raptor.properties do Presto.
prestosql-connector-redis	Altere os valores no arquivo SQL redis.properties do Presto.
prestosql-connector-redshift	Altere os valores no arquivo SQL redshift.properties do Presto.
prestosql-connector-tpch	Altere os valores no arquivo SQL tpch.properties do Presto.
prestosql-connector-tpcds	Altere os valores no arquivo SQL tpcds.properties do Presto.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.
ranger-kms-log4j	Altere os valores no arquivo kms-log4j.properties do Ranger. KMS
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.

Classificações	Descrição
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.

Classificações	Descrição
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

EMR Versão 6.0.1 da Amazon

Versões 6.0.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livvy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Presto](#), [Spark](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
AWS SDK para Java	1.11.828	1.11.828	1.11.711	1.11.711
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12.10	2.12.10	2.12.10	2.11.12
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

Notas da versão 6.0.1

Esta é uma versão para corrigir problemas com o Amazon EMR Scaling quando ele falha na escalabilidade ou redução de escala de um cluster com sucesso ou causa falhas no aplicativo.

Alterações, melhorias e problemas resolvidos

- Corrigido um problema em que as solicitações de escalabilidade falhavam em um cluster grande e altamente utilizado quando os daemons no cluster da EMR Amazon estavam executando atividades de verificação de integridade, como coletar o estado do nó e o estado do YARN nó. HDFS Isso estava acontecendo porque os daemons no cluster não conseguiam comunicar os dados do status de saúde de um nó aos componentes internos da Amazon. EMR
- Daemons EMR no cluster aprimorados para rastrear corretamente os estados dos nós quando os endereços IP são reutilizados para melhorar a confiabilidade durante as operações de escalabilidade.
- [SPARK-29683](#). Corrigido um problema em que ocorriam falhas de trabalho durante a redução da escala verticalmente do cluster, pois o Spark presumia que todos os nós disponíveis estavam na lista de negação.
- [YARN-9011](#). Corrigido um problema em que falhas de trabalho ocorriam devido a uma condição de corrida no YARN descomissionamento quando o cluster tentava aumentar ou diminuir a escala.

- Foi corrigido o problema com falhas de etapas ou tarefas durante a escalabilidade do cluster, garantindo que os estados dos nós fossem sempre consistentes entre os EMR daemons no cluster da Amazon e/. YARN HDFS
- Corrigido um problema em que operações de cluster, como redução de escala e envio de etapas, falhavam em EMR clusters da Amazon habilitados com a autenticação Kerberos. Isso ocorreu porque o daemon no EMR cluster da Amazon não renovou o tíquete Kerberos, que é necessário para se comunicar de forma segura com/em execução no nó primário. HDFS YARN
- As EMR versões mais recentes da Amazon corrigem o problema com um limite menor de “Máximo de arquivos abertos” em relação aos mais antigos AL2 na AmazonEMR. As EMR versões 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 e posteriores agora incluem uma correção permanente com uma configuração mais alta de “Máximo de arquivos abertos”.
- HTTPS agora está habilitado por padrão para os repositórios Amazon Linux. Se você estiver usando uma VPCE política do Amazon S3 para restringir o acesso a buckets específicos, deverá adicionar o novo bucket Amazon Linux ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` à sua política (`$region` substitua pela região onde está o endpoint). Para obter mais informações, consulte esse tópico nos fóruns de AWS discussão. [Anúncio: o Amazon Linux 2 agora oferece suporte à capacidade de uso HTTPS ao se conectar a repositórios de pacotes.](#)

Versões de componente 6.0.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
aws-sagemaker-spark-sdk	1.2.6	Amazon SageMaker Spark SDK
emr-ddb	4.14.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	3.0.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.5.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.14.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	1.5.0	EMRConector S3Select
emrfs	2.39.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas

Componente	Version (Versão)	Descrição
		coletadas pelo coletor de metadados do Ganglia.
hadoop-client	3.2.1-amzn-0.1	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	3.2.1-amzn-0.1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	3.2.1-amzn-0.1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	3.2.1-amzn-0.1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	3.2.1-amzn-0.1	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httfs-server	3.2.1-amzn-0.1	HTTPendpoint para HDFS operações.
hadoop-kms-server	3.2.1-amzn-0.1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	3.2.1-amzn-0.1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	3.2.1-amzn-0.1	YARNserviço para gerenciar contêineres em um nó individual.

Componente	Version (Versão)	Descrição
hadoop-yarn-resourcemanager	3.2.1-amzn-0.1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	3.2.1-amzn-0.1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	2.2.3	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	2.2.3	Serviço para atender a uma ou mais HBase regiões.
hbase-client	2.2.3	HBasecliente de linha de comando.
hbase-rest-server	2.2.3	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	2.2.3	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	3.1.2-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	3.1.2-amzn-0	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.

Componente	Version (Versão)	Descrição
hcatalog-webhcat-server	3.1.2-amzn-0	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	3.1.2-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	3.1.2-amzn-0	Cliente do Hive-hbase.
hive-metastore-server	3.1.2-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	3.1.2-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.5.0-incubating-amzn-1	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.
hudi-presto	0.5.0-incubating-amzn-1	Biblioteca de pacotes para executar o Presto com o Hudi.
hue-server	4.4.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	1.0.0	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.6.0-incubating	REST interface para interagir com o Apache Spark

Componente	Version (Versão)	Descrição
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mxnet	1.5.1	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mariadb-server	5.5.64+	Servidor de banco de dados MariaDB.
nvidia-cuda	9.2.88	Drivers NVIDIA e toolkit CUDA
oozie-client	5.1.0	O cliente da linha de comando do Oozie.
oozie-server	5.1.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	3.4.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	5.0.0- HBase -2,0	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	5.0.0- HBase -2,0	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.230	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.230	O serviço que executa partes de uma consulta.

Componente	Version (Versão)	Descrição
presto-client	0.230	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
r	3.4.3	O projeto R para computação estatística
spark-client	2.4.4	Os clientes da linha de comando do Spark.
spark-history-server	2.4.4	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.4.4	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.4.4	Bibliotecas do Apache Spark necessárias para escravos. YARN
tensorflow	1.14.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.2	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.41+	HTTP Servidor Apache.

Componente	Version (Versão)	Descrição
zeppelin-server	0.9.0- SNAPSHOT	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.14	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.14	ZooKeeper cliente de linha de comando.

Classificações de configuração 6.0.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-6.0.1

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
container-executor	Altere os valores no arquivo <code>container-executor.cfg</code> YARN do Hadoop.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.

Classificações	Descrição
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-env	Mude os valores no HDFS ambiente.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.

Classificações	Descrição
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive	Configurações EMR selecionadas pela Amazon para o Apache Hive.
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.

Classificações	Descrição
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.

Classificações	Descrição
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.

Classificações	Descrição
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.

Classificações	Descrição
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.
ranger-kms-log4j	Altere os valores no arquivo kms-log4j.properties do Ranger. KMS
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.
recordserver-env	Mude os valores no EMR RecordServer ambiente.
recordserver-conf	Altere os valores no arquivo EMR RecordServer erver.properties do.
recordserver-log4j	Altere os valores no EMR RecordServer arquivo log4j.properties.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.

Classificações	Descrição
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 6.0.0

Versões 6.0.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Oozie](#), [Phoenix](#), [Presto](#), [Spark](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
AWS SDK para Java	1.11.828	1.11.828	1.11.711	1.11.711
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.12.10	2.12.10	2.12.10	2.11.12
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
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

	emr-6.1.1	emr-6.1.0	emr-6.0.1	emr-6.0.0
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
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

Notas da versão 6.0.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 6.0.0 da Amazon.

Data da versão inicial: 10 de março de 2020

Aplicações compatíveis

- AWS SDK for Java versão 1.11.711
- Ganglia versão 3.7.2
- Hadoop versão 3.2.1
- HBase versão 2.2.3
- HCatalog versão 3.1.2
- Hive versão 3.1.2
- Hudi versão 0.5.0 incubadora
- Hue versão 4.4.0
- JupyterHub versão 1.0.0
- Livy versão 0.6.0
- MXNet versão 1.5.1
- Oozie versão 5.1.0
- Phoenix versão 5.0.0
- Presto versão 0.230
- Spark versão 2.4.4
- TensorFlow versão 1.14.0
- Zeppelin versão 0.9.0- SNAPSHOT
- Zookeeper versão 3.4.14
- Conectores e drivers: DynamoDB Connector 4.14.0

Note

Flink, Sqoop, Pig e Mahout não estão disponíveis na versão 6.0.0 da Amazon. EMR

Novos atributos

- YARN Docker Runtime Support — YARN aplicativos, como trabalhos do Spark, agora podem ser executados no contexto de um contêiner Docker. Isso permite que você defina

facilmente dependências em uma imagem do Docker sem a necessidade de instalar bibliotecas personalizadas em seu cluster da AmazonEMR. Para obter mais informações, consulte [Configurar a integração do Docker](#) e [executar aplicativos Spark com o Docker usando o Amazon 6.0.0](#). EMR

- LLAP Suporte do Hive - O Hive agora suporta o modo de LLAP execução para melhorar o desempenho das consultas. Para obter mais informações, consulte [Usando o Hive. LLAP](#)

Alterações, melhorias e problemas resolvidos

- Esta é uma versão para corrigir problemas com o Amazon EMR Scaling quando ele falha na escalabilidade ou redução de escala de um cluster com sucesso ou causa falhas no aplicativo.
- Corrigido um problema em que as solicitações de escalabilidade falhavam em um cluster grande e altamente utilizado quando os daemons no cluster da EMR Amazon estavam executando atividades de verificação de integridade, como coletar o estado do nó e o estado do YARN nó. HDFS Isso estava acontecendo porque os daemons no cluster não conseguiam comunicar os dados do status de saúde de um nó aos componentes internos da Amazon. EMR
- Daemons EMR no cluster aprimorados para rastrear corretamente os estados dos nós quando os endereços IP são reutilizados para melhorar a confiabilidade durante as operações de escalabilidade.
- [SPARK-29683](#). Corrigido um problema em que ocorriam falhas de trabalho durante a redução da escala verticalmente do cluster, pois o Spark presumia que todos os nós disponíveis estavam na lista de negação.
- [YARN-9011](#). Corrigido um problema em que falhas de trabalho ocorriam devido a uma condição de corrida no YARN descomissionamento quando o cluster tentava aumentar ou diminuir a escala.
- Foi corrigido o problema com falhas de etapas ou tarefas durante a escalabilidade do cluster, garantindo que os estados dos nós fossem sempre consistentes entre os EMR daemons no cluster da Amazon e/. YARN HDFS
- Corrigido um problema em que operações de cluster, como redução de escala e envio de etapas, falhavam em EMR clusters da Amazon habilitados com a autenticação Kerberos. Isso ocorreu porque o daemon no EMR cluster da Amazon não renovou o tíquete Kerberos, que é necessário para se comunicar de forma segura com/em execução no nó primário. HDFS YARN
- As EMR versões mais recentes da Amazon corrigem o problema com um limite menor de “Máximo de arquivos abertos” em relação aos mais antigos AL2 na AmazonEMR. As EMR versões 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 e posteriores agora incluem uma correção permanente com uma configuração mais alta de “Máximo de arquivos abertos”.
- Amazon Linux

- O Amazon Linux 2 é o sistema operacional da série de lançamento EMR 6.x.
- `systemd` é usado para gerenciamento de serviços em vez do `in`Amazon Linux 1 `upstart` usado.
- Kit de desenvolvimento Java (JDK)
 - Corretto JDK 8 é o padrão JDK para a série de lançamento EMR 6.x.
- Scala
 - O Scala 2.12 é usado com o Apache Spark e com o Apache Livy.
- Python 3
 - O Python 3 agora é a versão padrão do Python em. EMR
- YARN rótulos de nós
 - A partir da série de lançamento Amazon EMR 6.x, o recurso de rótulos de YARN nós está desativado por padrão. Os principais processos do aplicativo podem ser executados tanto nos nós core como nos nós de tarefa por padrão. Você pode ativar o recurso de rótulos de YARN nós configurando as seguintes propriedades: `yarn.node-labels.enabled` e `yarn.node-labels.am.default-node-label-expression` Para obter mais informações, consulte [Noções básicas sobre nós de tarefa, centrais e primários](#).

Problemas conhecidos

- Limite inferior de “Máximo de arquivos abertos” em versões mais antigas AL2 [corrigido em versões mais recentes]. EMR Lançamentos da Amazon: `emr-5.30.x`, `emr-5.31.0`, `emr-5.32.0`, `emr-6.0.0`, `emr-6.1.0` e `emr-6.2.0` são baseados em versões mais antigas do Linux of Amazon 2 (AL2), que têm uma configuração de limite inferior para “Máximo de arquivos abertos” quando os clusters da Amazon são criados com o padrão. EMR AMI As EMR versões 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 e posteriores incluem uma correção permanente com uma configuração maior de “Máximo de arquivos abertos”. Versões com o limite inferior de arquivos abertos causam o erro “Muitos arquivos abertos” ao ser enviado um trabalho do Spark. Nas versões afetadas, o padrão da Amazon AMI tem uma configuração EMR padrão de `ulimit` de 4096 para “Máximo de arquivos abertos”, que é inferior ao limite de 65536 arquivos no Linux 2. latest Amazon AMI A configuração inferior de `ulimit` para “Máximo de arquivos abertos” causa falhas em trabalhos do Spark quando o driver e o executor do Spark tentam abrir mais de 4096 arquivos. Para corrigir o problema, a Amazon EMR tem um script de ação de bootstrap (BA) que ajusta a configuração `ulimit` na criação do cluster.

Se você estiver usando uma EMR versão mais antiga da Amazon que não tem a solução permanente para esse problema, a solução alternativa a seguir permite que você defina explicitamente o ulimit do controlador de instância para um máximo de 65536 arquivos.

Defina explicitamente um ulimit na linha de comando

1. Edite `/etc/systemd/system/instance-controller.service` para adicionar os seguintes parâmetros à seção Serviço.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Reiniciar InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Defina um ulimit usando a ação de bootstrap (BA)

Você também pode usar um script de ação de bootstrap (BA) para configurar o ulimit instance-controller para 65536 arquivos na criação do 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
```

- O shell interativo do Spark PySpark, incluindo o SparkR e o spark-shell, não é compatível com o uso do Docker com bibliotecas adicionais.
- Para usar o Python 3 com a EMR versão 6.0.0 da Amazon, você deve adicionar a. `PATH yarn.nodemanager.env-whitelist`
- A funcionalidade Live Long and Process (LLAP) não é suportada quando você usa o AWS Glue Data Catalog como metastore do Hive.
- Ao usar o Amazon EMR 6.0.0 com a integração do Spark e do Docker, você precisa configurar as instâncias em seu cluster com o mesmo tipo de instância e a mesma quantidade de EBS volumes para evitar falhas ao enviar um trabalho do Spark com o tempo de execução do Docker.
- [No Amazon EMR 6.0.0, HBase no Amazon S3, o modo de armazenamento é afetado pelo HBASE problema -24286](#). HBaseo master não pode inicializar quando o cluster é criado usando dados existentes do S3.
- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSH como hadoop usuário do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do hadoop.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo keytab está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas

definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

Versões de componente 6.0.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.2.6	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.14.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	3.0.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.5.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.14.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.

Componente	Version (Versão)	Descrição
emr-s3-select	1.5.0	EMRConector S3Select
emrfs	2.39.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	3.2.1-amzn-0	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	3.2.1-amzn-0	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	3.2.1-amzn-0	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	3.2.1-amzn-0	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.

Componente	Version (Versão)	Descrição
hadoop-hdfs-journalnode	3.2.1-amzn-0	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httfs-server	3.2.1-amzn-0	HTTPendpoint para HDFS operações.
hadoop-kms-server	3.2.1-amzn-0	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	3.2.1-amzn-0	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	3.2.1-amzn-0	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	3.2.1-amzn-0	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	3.2.1-amzn-0	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	2.2.3	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	2.2.3	Serviço para atender a uma ou mais HBase regiões.

Componente	Version (Versão)	Descrição
hbase-client	2.2.3	HBase cliente de linha de comando.
hbase-rest-server	2.2.3	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	2.2.3	Serviço que fornece um endpoint Thrift para HBase
hcatalog-client	3.1.2-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	3.1.2-amzn-0	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	3.1.2-amzn-0	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	3.1.2-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	3.1.2-amzn-0	Cliente do Hive-hbase.
hive-metastore-server	3.1.2-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL

Componente	Version (Versão)	Descrição
hive-server2	3.1.2-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.5.0-incubating-amzn-1	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.
hudi-presto	0.5.0-incubating-amzn-1	Biblioteca de pacotes para executar o Presto com o Hudi.
hue-server	4.4.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	1.0.0	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.6.0-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mxnet	1.5.1	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mariadb-server	5.5.64+	Servidor de banco de dados MariaDB.
nvidia-cuda	9.2.88	Drivers NVIDIA e toolkit CUDA
oozie-client	5.1.0	O cliente da linha de comando do Oozie.

Componente	Version (Versão)	Descrição
oozie-server	5.1.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	3.4.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	5.0.0- HBase -2,0	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	5.0.0- HBase -2,0	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.230	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.230	O serviço que executa partes de uma consulta.
presto-client	0.230	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
r	3.4.3	O projeto R para computação estatística
spark-client	2.4.4	Os clientes da linha de comando do Spark.

Componente	Version (Versão)	Descrição
spark-history-server	2.4.4	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.4.4	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.4.4	Bibliotecas do Apache Spark necessárias para escravos. YARN
tensorflow	1.14.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.2	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.41+	HTTP Servidor Apache.
zeppelin-server	0.9.0- SNAPSHOT	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.14	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.14	ZooKeeper cliente de linha de comando.

Classificações de configuração 6.0.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-6.0.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
container-executor	Altere os valores no arquivo <code>container-executor.cfg</code> YARN do Hadoop.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.

Classificações	Descrição
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-env	Mude os valores no HDFS ambiente.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive	Configurações EMR selecionadas pela Amazon para o Apache Hive.

Classificações	Descrição
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.

Classificações	Descrição
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.

Classificações	Descrição
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.

Classificações	Descrição
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.

Classificações	Descrição
ranger-kms-log4j	Altere os valores no arquivo kms-log4j.properties do Ranger. KMS
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.
recordserver-env	Mude os valores no EMR RecordServer ambiente.
recordserver-conf	Altere os valores no arquivo EMR RecordServer erver.properties do.
recordserver-log4j	Altere os valores no EMR RecordServer arquivo log4j.properties.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.

Classificações	Descrição
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Versões de lançamento do Amazon EMR 5.x

Esta seção contém versões de aplicativos, notas de lançamento, versões de componentes e classificações de configuração disponíveis em cada versão de lançamento do Amazon EMR 5.x.

Ao iniciar um cluster, você pode escolher entre várias versões da AmazonEMR. Isso permite que você teste e use versões de aplicações que atendem aos seus requisitos de compatibilidade. Você especifica o número da versão com o rótulo da versão. Os rótulos de versão estão no formato `emr-x.x.x`. Por exemplo, `emr-7.2.0`.

Novos EMR lançamentos da Amazon são disponibilizados em diferentes regiões por um período de vários dias, começando com a primeira região na data de lançamento inicial. A versão mais recente pode não estar disponível em sua região durante esse período.

Para obter uma tabela abrangente das versões do aplicativo em cada versão EMR 5.x da Amazon, consulte [Versões de aplicações nas versões 5.x do Amazon EMR](#).

Tópicos

- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Amazon EMR versão 5.36.2](#)
- [Amazon EMR versão 5.36.1](#)
- [Amazon EMR versão 5.36.0](#)
- [Amazon EMR versão 5.35.0](#)
- [Amazon EMR versão 5.34.0](#)

- [Amazon EMR versão 5.33.1](#)
- [Amazon EMR versão 5.33.0](#)
- [Amazon EMR versão 5.32.1](#)
- [Amazon EMR versão 5.32.0](#)
- [Amazon EMR versão 5.31.1](#)
- [Amazon EMR versão 5.31.0](#)
- [Amazon EMR versão 5.30.2](#)
- [Amazon EMR versão 5.30.1](#)
- [Amazon EMR versão 5.30.0](#)
- [Amazon EMR versão 5.29.0](#)
- [Amazon EMR versão 5.28.1](#)
- [Amazon EMR versão 5.28.0](#)
- [Amazon EMR versão 5.27.1](#)
- [Amazon EMR versão 5.27.0](#)
- [Amazon EMR versão 5.26.0](#)
- [Amazon EMR versão 5.25.0](#)
- [Amazon EMR versão 5.24.1](#)
- [Amazon EMR versão 5.24.0](#)
- [Amazon EMR versão 5.23.1](#)
- [Amazon EMR versão 5.23.0](#)
- [Amazon EMR versão 5.22.0](#)
- [Amazon EMR versão 5.21.2](#)
- [Amazon EMR versão 5.21.1](#)
- [Amazon EMR versão 5.21.0](#)
- [Amazon EMR versão 5.20.1](#)
- [Amazon EMR versão 5.20.0](#)
- [Amazon EMR versão 5.19.1](#)
- [Amazon EMR versão 5.19.0](#)
- [Amazon EMR versão 5.18.1](#)

- [Amazon EMR versão 5.18.0](#)
- [Amazon EMR versão 5.17.2](#)
- [Amazon EMR versão 5.17.1](#)
- [Amazon EMR versão 5.17.0](#)
- [Amazon EMR versão 5.16.1](#)
- [Amazon EMR versão 5.16.0](#)
- [Amazon EMR versão 5.15.1](#)
- [Amazon EMR versão 5.15.0](#)
- [Amazon EMR versão 5.14.2](#)
- [Amazon EMR versão 5.14.1](#)
- [Amazon EMR versão 5.14.0](#)
- [Amazon EMR versão 5.13.1](#)
- [Amazon EMR versão 5.13.0](#)
- [Amazon EMR versão 5.12.3](#)
- [Amazon EMR versão 5.12.2](#)
- [Amazon EMR versão 5.12.1](#)
- [Amazon EMR versão 5.12.0](#)
- [Amazon EMR versão 5.11.4](#)
- [Amazon EMR versão 5.11.3](#)
- [Amazon EMR versão 5.11.2](#)
- [Amazon EMR versão 5.11.1](#)
- [Amazon EMR versão 5.11.0](#)
- [Amazon EMR versão 5.10.1](#)
- [Amazon EMR versão 5.10.0](#)
- [Amazon EMR versão 5.9.1](#)
- [Amazon EMR versão 5.9.0](#)
- [Amazon EMR versão 5.8.3](#)
- [Amazon EMR versão 5.8.2](#)
- [Amazon EMR versão 5.8.1](#)

- [Amazon EMR versão 5.8.0](#)
- [Amazon EMR versão 5.7.1](#)
- [Amazon EMR versão 5.7.0](#)
- [Amazon EMR versão 5.6.1](#)
- [Amazon EMR versão 5.6.0](#)
- [Amazon EMR versão 5.5.4](#)
- [Amazon EMR versão 5.5.3](#)
- [Amazon EMR versão 5.5.2](#)
- [Amazon EMR versão 5.5.1](#)
- [Amazon EMR versão 5.5.0](#)
- [Amazon EMR versão 5.4.1](#)
- [Amazon EMR versão 5.4.0](#)
- [Amazon EMR versão 5.3.2](#)
- [Amazon EMR versão 5.3.1](#)
- [Amazon EMR versão 5.3.0](#)
- [Amazon EMR versão 5.2.3](#)
- [Amazon EMR versão 5.2.2](#)
- [Amazon EMR versão 5.2.1](#)
- [Amazon EMR versão 5.2.0](#)
- [Amazon EMR versão 5.1.1](#)
- [Amazon EMR versão 5.1.0](#)
- [Amazon EMR versão 5.0.3](#)
- [Amazon EMR versão 5.0.2](#)
- [Amazon EMR versão 5.0.1](#)
- [Amazon EMR versão 5.0.0](#)

Versões de aplicações nas versões 5.x do Amazon EMR

Para obter uma tabela abrangente que liste as versões das aplicações disponíveis em cada versão 5.x do Amazon EMR, abra as [versões de aplicações nas versões 5.x do Amazon EMR](#) no seu navegador.

Amazon EMR versão 5.36.2

5.36.2 versões do aplicativo

As seguintes aplicações são compatíveis com esta versão: [Delta](#), [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.36.2	emr-5.36.1	emr-5.36.0	emr-5.35.0
AWS SDK para 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	2.11.12	2.11.12	2.11.12	2.11.12
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

Notas de lançamento do 5.36.2

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.36.2 da Amazon. As alterações são relativas ao 5.36.1. Para obter informações sobre o cronograma da versão, consulte o [log de alterações](#).

Alterações, melhorias e problemas resolvidos

- Esta versão melhora a lógica de redução do cluster para que a Amazon EMR não reduza os nós principais abaixo da configuração do fator de HDFS replicação do cluster. Essa melhoria atende aos requisitos de redundância de dados e reduz a chance de uma operação de escalabilidade parar.
- Esta versão adiciona um novo mecanismo de repetição ao fluxo de trabalho de escalabilidade de cluster para aquela execução Presto ou Trino. Essa melhoria reduz o risco de o redimensionamento do cluster ser executado indefinidamente devido a uma única falha na operação de redimensionamento. Ela também aprimora a utilização dos clusters, porque seu cluster aumenta e reduz a escala verticalmente com mais rapidez.
- Corrige um problema em que as operações de redução de escala do cluster podem parar enquanto a Amazon EMR descomissiona normalmente um nó principal e ele fica insalubre antes de ser totalmente desativado.
- Melhora a estabilidade de um nó em um cluster de alta disponibilidade com vários nós primários quando a Amazon EMR reinicia um único nó.
- Otimiza o gerenciamento de registros com a Amazon em EMR execução na AmazonEC2. Como resultado, é possível ver uma pequena redução nos custos de armazenamento dos logs do cluster.
- Melhora o gerenciamento dos arquivos de log de ZooKeeper transações que são mantidos nos nós primários para minimizar os cenários em que os arquivos de log ultrapassam os limites e interrompem as operações do cluster.

- Corrige um bug raro que pode fazer com que um cluster de alta disponibilidade com vários nós primários falhe por não conseguir se comunicar com o ResourceManager Yarn.
- Quando você inicia um cluster com a versão de patch mais recente do Amazon EMR 5.36 ou superior, 6.6 ou superior ou 7.0 ou superior, a Amazon EMR usa a versão mais recente do Amazon Linux 2023 ou Amazon Linux 2 para a Amazon padrão. EMR AMI Para obter mais informações, consulte [Usando o Amazon Linux padrão AMI para a Amazon EMR](#).

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 709.1	4.14.348	23 de julho de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(Central), (Oeste dos EUA), (Leste dos EUA), AWS GovCloud China (Pequim), China AWS GovCloud (Ningxia), Ásia Pacífico (Hyderabad), Oriente Médio (), Europa (EspanhaUAE), Europa (Zurique), Ásia-Pacífico (Melbourne), Israel (Tel Aviv), Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 503.0	4.14.343	xxxxxx, 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

Versões do componente 5.36.2

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.16.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.16.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.5.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-notebook-env</code>	1.5.0	Ambiente Conda para bloco de anotações do emr, que inclui o jupyter enterprise gateway
<code>emr-s3-dist-cp</code>	2.21.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.

Componente	Version (Versão)	Descrição
emr-s3-select	1.7.0	EMRConector S3Select
emrfs	2.51.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.14.2	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
flink-jobmanager-config	1.14.2	Gerenciando recursos em EMR nós para o Apache JobManager Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.10.1-amzn-4	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.10.1-amzn-4	HDFSserviço em nível de nó para armazenar blocos.

Componente	Version (Versão)	Descrição
hadoop-hdfs-library	2.10.1-amzn-4	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.10.1-amzn-4	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	2.10.1-amzn-4	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-https-server	2.10.1-amzn-4	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.10.1-amzn-4	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.10.1-amzn-4	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.10.1-amzn-4	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.10.1-amzn-4	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.10.1-amzn-4	Serviço para recuperar informações atuais e históricas para YARN aplicativos.

Componente	Version (Versão)	Descrição
hbase-hmaster	1.4.13	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.13	Serviço para atender uma ou mais HBase regiões.
hbase-client	1.4.13	HBase cliente de linha de comando.
hbase-rest-server	1.4.13	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.4.13	Serviço que fornece um endpoint Thrift para HBase.
hcatalog-client	2,3.9-amzn-2	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2,3.9-amzn-2	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2,3.9-amzn-2	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2,3.9-amzn-2	O cliente da linha de comando do Hive.
hive-hbase	2,3.9-amzn-2	Cliente do Hive-hbase.

Componente	Version (Versão)	Descrição
hive-metastore-server	2,3.9-amzn-2	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2,3.9-amzn-2	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.10.1-amzn-1	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.
hudi-spark	0.10.1-amzn-1	Biblioteca de pacotes para executar o Spark com o Hudi.
hudi-presto	0.10.1-amzn-1	Biblioteca de pacotes para executar o Presto com o Hudi.
hue-server	4.10.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	1.4.1	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.7.1-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.13.1	nginx [engine x] é HTTP um servidor proxy reverso
mahout-client	0.13.0	A biblioteca de Machine Learning.

Componente	Version (Versão)	Descrição
mxnet	1.8.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mariadb-server	5.5.68	Meu servidor SQL de banco de dados.
nvidia-cuda	11.0.194	Drivers NVIDIA e toolkit CUDA
oozie-client	5.2.1	O cliente da linha de comando do Oozie.
oozie-server	5.2.1	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	4.5.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.3- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.3- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.267-amzn-1	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.267-amzn-1	O serviço que executa partes de uma consulta.

Componente	Version (Versão)	Descrição
presto-client	0.267-amzn-1	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	4.0.2	O projeto R para computação estatística
ranger-kms-server	1.2.0	Sistema de gerenciamento de chaves Apache Ranger
spark-client	2.4.8-amzn-2	Os clientes da linha de comando do Spark.
spark-history-server	2.4.8-amzn-2	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.4.8-amzn-2	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.4.8-amzn-2	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.

Componente	Version (Versão)	Descrição
tensorflow	2.4.1	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.2	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.10.0	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.14	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.14	ZooKeeper cliente de linha de comando.

5.36.2 classificações de configuração

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Ocorrem ações de reconfiguração quando você especifica uma configuração para grupos de instâncias em um cluster em execução. A Amazon EMR só inicia ações de reconfiguração para as classificações que você modifica. Para ter mais informações, consulte [Reconfigurar um grupo de instâncias de um cluster em execução](#).

classificações do emr-5.36.2

Classificações	Descrição	Ações de reconfiguração
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.	Restarts the Resource Manager service.
container-executor	Altere os valores no arquivo container-executor.cfg YARN do Hadoop.	Not available.
container-log4j	Altere os valores no arquivo container-log4j.properties YARN do Hadoop.	Not available.
core-site	Alterar os valores no arquivo core-site.xml do 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	Alterar as configurações relacionadas ao docker.	Not available.
emrfs-site	Altere EMRFS as configurações.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

Classificações	Descrição	Ações de reconfiguração
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Alterar as configurações do flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Alterar as configurações de log4j.properties no Flink.	Restarts Flink history server.
flink-log4j-session	Alterar as configurações de log4j-yarn-session.properties no Flink para sessão do Kubernetes/Yarn.	Not available.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.	Restarts Flink history server.

Classificações	Descrição	Ações de reconfiguração
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do 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.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do 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 MapReduceHistoryServer.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop	Not available.
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop	Not available.

Classificações	Descrição	Ações de reconfiguração
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Mude os valores no HBase ambiente.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.	Not available.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.	Should not be reconfigured.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Classificações	Descrição	Ações de reconfiguração
hcatalog-env	Mude os valores no HCatalog ambiente.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.	Not available.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.	Not available.
hive-env	Alterar os valores no ambiente do Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log 4j2.properties do Hive.	Restarts HiveServer2 and HiveMetastore.

Classificações	Descrição	Ações de reconfiguração
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log 4j2.properties do Hive.	Not available.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.	Not available.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.	Not available.
hue-ini	Alterar os valores no arquivo ini do Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Mude os valores no HTTPFS ambiente.	Restarts Hadoop Httpfs service.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.	Not available.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.	Not available.

Classificações	Descrição	Ações de reconfiguração
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Altere os valores no ambiente do Hudi.	Not available.
hudi-defaults	Alterar os valores no arquivo hudi-defaults.conf do Hudi.	Not available.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.	Not available.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.	Not available.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.	Not available.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.	Not available.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.	Restarts Livy Server.
livy-env	Alterar os valores no ambiente do Livy.	Restarts Livy Server.
livy-log4j	Alterar as configurações de log4j.properties no Livy.	Restarts Livy Server.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.

Classificações	Descrição	Ações de reconfiguração
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Alterar os valores no ambiente do Oozie.	Restarts Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.	Restarts Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.	Not available.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.	Not available.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.	Not available.
pig-env	Alterar os valores no ambiente do Pig.	Not available.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.	Restarts Oozie.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-log	Alterar os valores no arquivo log.properties do Presto.	Restarts Presto-Server.
presto-config	Alterar os valores no arquivo config.properties do Presto.	Restarts Presto-Server.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.	Not available.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.	Restarts Presto-Server.
presto-node	Altere os valores no arquivo node.properties do Presto.	Not available.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.	Not available.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.	Not available.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.	Restarts Presto-Server.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.	Not available.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.	Not available.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.	Not available.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.	Not available.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.	Not available.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.	Not available.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.	Not available.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.	Not available.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.	Not available.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.	Not available.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.	Not available.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.	Restarts Ranger KMS Server.

Classificações	Descrição	Ações de reconfiguração
ranger-kms-env	Altere os valores no KMS ambiente Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Altere os valores no arquivo kms-log4j.properties do Ranger. KMS	Not available.
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.	Not available.
recordserver-env	Mude os valores no EMR RecordServer ambiente.	Restarts EMR record server.
recordserver-conf	Altere os valores no arquivo EMR RecordServer erver.properties do.	Restarts EMR record server.
recordserver-log4j	Altere os valores no EMR RecordServer arquivo log4j.properties.	Restarts EMR record server.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Alterar os valores no ambiente do Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.	Not available.

Classificações	Descrição	Ações de reconfiguração
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Alterar os valores no ambiente do Sqoop.	Not available.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.	Not available.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.	Not available.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.	Restarts Oozie.
yarn-env	Mude os valores no YARN ambiente.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.	Restarts Zeppelin.

Classificações	Descrição	Ações de reconfiguração
zeppelin-site	Alterar as definições da configuração em zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.	Restarts Zookeeper server.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.	Restarts Zookeeper server.

5.36.2 registro de alterações

Registro de alterações da versão 5.36.2 e das notas de lançamento

Data	Evento	Descrição
2024-05-28	Implantação concluída	Amazon EMR 5.36.2 totalmente implantado em todas as regiões suportadas
2024-05-28	Publicação de documentos	Notas de lançamento do Amazon EMR 5.36.2 publicadas pela primeira vez
2024-05-20	Lançamento inicial	Amazon EMR 5.36.2 foi implantado pela primeira vez em regiões comerciais limitadas

Amazon EMR versão 5.36.1

Versões 5.36.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.36.1	emr-5.36.0	emr-5.35.0	emr-5.34.0
AWS SDK para Java	1.12.206	1.12.206	1.12.159	1.11.970
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.11.12	2.11.12	2.11.12	2.11.12
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

	emr-5.36.1	emr-5.36.0	emr-5.35.0	emr-5.34.0
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
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

	emr-5.36.1	emr-5.36.0	emr-5.35.0	emr-5.34.0
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

Notas da versão 5.36.1

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.36.1 da Amazon. As alterações são referentes à versão 5.36.0. Para obter informações sobre o cronograma da versão, consulte o [log de alterações](#).

Alterações, melhorias e problemas resolvidos

- A EMR versão 5.36.1 da Amazon adiciona suporte ao arquivamento de registros no Amazon S3 durante a redução da escala do cluster. Nas versões 5.x anteriores, só era possível arquivar arquivos de log no Amazon S3 durante o encerramento do cluster. Essa melhoria garante que os arquivos de log gerados no cluster persistam no Amazon S3 mesmo após o encerramento do nó. Para obter mais informações, consulte [Configurar registro em log e depuração do cluster](#).
- A versão 5.36.1 aprimora o daemon de gerenciamento de logs no cluster para monitorar pastas de log adicionais em seu cluster. EMR Essa melhoria minimiza os cenários de utilização excessiva do disco.
- A versão 5.36.1 reinicia automaticamente o daemon de gerenciamento de logs no cluster quando ele é interrompido. Essa melhoria reduz o risco de os nós parecerem não íntegros devido à utilização excessiva do disco.
- A versão 5.36.1 corrige um problema em que os EMR daemons da Amazon no nó primário manteriam metadados obsoletos para instâncias encerradas no cluster. Manter dados obsoletos pode fazer com que o uso do cluster CPU e da memória cresça sem limites e, por fim, causar falhas no cluster.
- Para clusters que são lançados com vários nós primários, a versão 5.36.1 corrige um problema em que uma falha de EC2 hardware da Amazon em um dos nós primários pode causar a falha de um segundo nó primário e tornar seu cluster instável.

- Para clusters configurados com criptografia em trânsito, o Ajuste de escala gerenciado agora reconhece dados embaralhados do Spark. dados embaralhados do Spark são dados que o Spark redistribui entre partições para executar operações específicas. Durante a redução da escala verticalmente, o escalonamento gerenciado ignora as instâncias com dados embaralhados. Isso evita novas tentativas de trabalho e recálculos, que são onerosos em termos de preço e performance. Para obter mais informações sobre operações de shuffle, consulte o [Guia de programação do Spark](#).
- Quando você inicia um cluster com a versão de patch mais recente do Amazon EMR 5.36 ou superior, 6.6 ou superior ou 7.0 ou superior, a Amazon EMR usa a versão mais recente do Amazon Linux 2023 ou Amazon Linux 2 para a Amazon padrão. EMR AMI Para obter mais informações, consulte [Usando o Amazon Linux padrão AMI para a Amazon EMR](#).

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 709.1	4.14.348	23 de julho de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), AWS GovCloud China (Pequim), China AWS GovCloud (Ningxia), Ásia Pacífico (Hyderabad), Oriente Médio (), Europa (EspanhaUAE), Europa (Zurique), Ásia-Pacífico (Melbourne), Israel (Tel Aviv), Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 223.0	4.14.336	8 de março de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 131.0	4.14.336	14 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 124.0	4.14.336	7 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 109.0	4.14.334	24 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 218.0	4.14.330	2 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 206.0	4.14.330	22 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 116.0	4.14.328	11 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 101.0	4.14.327	16 de novembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 020.1	4.14.326	7 de novembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 012.1	4.14.326	26 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 926.0	4.14.322	19 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 906.0	4.14.322	4 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 822.0	4.14.322	30 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 808.0	4.14.320	24 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 727.0	4.14.320	14 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 719.0	4.14.320	2 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)
2.0.2023 628.0	4.14.318	12 de julho de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 612.0	4.14.314	23 de junho de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 404.1	4.14.311	18 de abril de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central)

Versões de componente 5.36.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são

a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado myapp-component com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como 2.2-amzn-2.

Componente	Version (Versão)	Descrição
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK
emr-ddb	4.16.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.16.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.5.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-notebook-env	1.5.0	Ambiente Conda para bloco de anotações do emr, que inclui o jupyter enterprise gateway
emr-s3-dist-cp	2.21.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	1.7.0	EMRConector S3Select
emrfs	2.51.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
flink-client	1.14.2	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
flink-jobmanager-config	1.14.2	Gerenciando recursos em EMR nós para o Apache JobManager Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.10.1-amzn-4	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.10.1-amzn-4	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.10.1-amzn-4	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.10.1-amzn-4	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.

Componente	Version (Versão)	Descrição
hadoop-hdfs-journalnode	2.10.1-amzn-4	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httfs-server	2.10.1-amzn-4	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.10.1-amzn-4	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.10.1-amzn-4	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.10.1-amzn-4	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.10.1-amzn-4	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.10.1-amzn-4	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.13	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.13	Serviço para atender uma ou mais HBase regiões.

Componente	Version (Versão)	Descrição
hbase-client	1.4.13	HBase cliente de linha de comando.
hbase-rest-server	1.4.13	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.4.13	Serviço que fornece um endpoint Thrift para HBase
hcatalog-client	2,3.9-amzn-2	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2,3.9-amzn-2	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2,3.9-amzn-2	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2,3.9-amzn-2	O cliente da linha de comando do Hive.
hive-hbase	2,3.9-amzn-2	Cliente do Hive-hbase.
hive-metastore-server	2,3.9-amzn-2	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL

Componente	Version (Versão)	Descrição
hive-server2	2,3.9-amzn-2	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.10.1-amzn-1	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.
hudi-spark	0.10.1-amzn-1	Biblioteca de pacotes para executar o Spark com o Hudi.
hudi-presto	0.10.1-amzn-1	Biblioteca de pacotes para executar o Presto com o Hudi.
hue-server	4.10.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	1.4.1	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.7.1-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.13.1	nginx [engine x] é HTTP um servidor proxy reverso
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.8.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.

Componente	Version (Versão)	Descrição
mariadb-server	5.5.68	Meu servidor SQL de banco de dados.
nvidia-cuda	11.0.194	Drivers NVIDIA e toolkit CUDA
oozie-client	5.2.1	O cliente da linha de comando do Oozie.
oozie-server	5.2.1	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	4.5.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.3- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.3- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.267-amzn-1	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.267-amzn-1	O serviço que executa partes de uma consulta.
presto-client	0.267-amzn-1	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.

Componente	Version (Versão)	Descrição
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	4.0.2	O projeto R para computação estatística
ranger-kms-server	1.2.0	Sistema de gerenciamento de chaves Apache Ranger
spark-client	2.4.8-amzn-2	Os clientes da linha de comando do Spark.
spark-history-server	2.4.8-amzn-2	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.4.8-amzn-2	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.4.8-amzn-2	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	2.4.1	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.2	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.

Componente	Version (Versão)	Descrição
zeppelin-server	0.10.0	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.14	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.14	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.36.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Ocorrem ações de reconfiguração quando você especifica uma configuração para grupos de instâncias em um cluster em execução. A Amazon EMR só inicia ações de reconfiguração para as classificações que você modifica. Para ter mais informações, consulte [Reconfigurar um grupo de instâncias de um cluster em execução](#).

Classificações do emr-5.36.1

Classificações	Descrição	Ações de reconfiguração
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.	Restarts the Resource Manager service.
container-executor	Altere os valores no arquivo <code>container-executor.cfg</code> YARN do Hadoop.	Not available.

Classificações	Descrição	Ações de reconfiguração
container-log4j	Altere os valores no arquivo container-log4j.properties YARN do Hadoop.	Not available.
core-site	Alterar os valores no arquivo core-site.xml do 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	Alterar as configurações relacionadas ao docker.	Not available.

Classificações	Descrição	Ações de reconfiguração
emrfs-site	Altere EMRFS as configurações.	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	Alterar as configurações do <code>flink-conf.yaml</code> .	Restarts Flink history server.
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.	Restarts Flink history server.
flink-log4j-session	Alterar as configurações de <code>log4j-yarn-session.properties</code> no Flink para sessão do Kubernetes/Yarn.	Not available.
flink-log4j-cli	Alterar as configurações de <code>log4j-cli.properties</code> no Flink.	Restarts Flink history server.

Classificações	Descrição	Ações de reconfiguração
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do 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.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do 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 MapReduceHistoryServer.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop	Not available.
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop	Not available.

Classificações	Descrição	Ações de reconfiguração
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Mude os valores no HBase ambiente.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.	Not available.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.	Should not be reconfigured.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Classificações	Descrição	Ações de reconfiguração
hcatalog-env	Mude os valores no HCatalog ambiente.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.	Not available.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.	Not available.
hive-env	Alterar os valores no ambiente do Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log 4j2.properties do Hive.	Restarts HiveServer2 and HiveMetastore.

Classificações	Descrição	Ações de reconfiguração
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log 4j2.properties do Hive.	Not available.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.	Not available.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.	Not available.
hue-ini	Alterar os valores no arquivo ini do Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Mude os valores no HTTPFS ambiente.	Restarts Hadoop Httpfs service.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.	Not available.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.	Not available.

Classificações	Descrição	Ações de reconfiguração
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Altere os valores no ambiente do Hudi.	Not available.
hudi-defaults	Alterar os valores no arquivo hudi-defaults.conf do Hudi.	Not available.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.	Not available.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.	Not available.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.	Not available.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.	Not available.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.	Restarts Livy Server.
livy-env	Alterar os valores no ambiente do Livy.	Restarts Livy Server.
livy-log4j	Alterar as configurações de log4j.properties no Livy.	Restarts Livy Server.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.

Classificações	Descrição	Ações de reconfiguração
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Alterar os valores no ambiente do Oozie.	Restarts Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.	Restarts Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.	Not available.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.	Not available.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.	Not available.
pig-env	Alterar os valores no ambiente do Pig.	Not available.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.	Restarts Oozie.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-log	Alterar os valores no arquivo log.properties do Presto.	Restarts Presto-Server.
presto-config	Alterar os valores no arquivo config.properties do Presto.	Restarts Presto-Server.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.	Not available.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.	Restarts Presto-Server.
presto-node	Altere os valores no arquivo node.properties do Presto.	Not available.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.	Not available.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.	Not available.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.	Restarts Presto-Server.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.	Not available.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.	Not available.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.	Not available.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.	Not available.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.	Not available.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.	Not available.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.	Not available.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.	Not available.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.	Not available.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.	Not available.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.	Not available.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.	Restarts Ranger KMS Server.

Classificações	Descrição	Ações de reconfiguração
ranger-kms-env	Altere os valores no KMS ambiente Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Altere os valores no arquivo kms-log4j.properties do Ranger. KMS	Not available.
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.	Not available.
recordserver-env	Mude os valores no EMR RecordServer ambiente.	Restarts EMR record server.
recordserver-conf	Altere os valores no arquivo EMR RecordServer erver.properties do.	Restarts EMR record server.
recordserver-log4j	Altere os valores no EMR RecordServer arquivo log4j.properties.	Restarts EMR record server.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Alterar os valores no ambiente do Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.	Not available.

Classificações	Descrição	Ações de reconfiguração
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Alterar os valores no ambiente do Sqoop.	Not available.
sqoop-oraoop-site	Altere os valores no arquivo ooraoop-site.xml OraOop do Sqoop.	Not available.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.	Not available.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.	Restarts Oozie.
yarn-env	Mude os valores no YARN ambiente.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.	Restarts Zeppelin.

Classificações	Descrição	Ações de reconfiguração
zeppelin-site	Alterar as definições da configuração em zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.	Restarts Zookeeper server.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.	Restarts Zookeeper server.

Log de alterações 5.36.1

Log de alterações da versão 5.36.1 e das notas da versão

Data	Evento	Descrição
2023-07-26	Atualizar	Novos rótulos de versão 2.0.20230612.0 e 2.0.20230628.0 do sistema operacional.
2023-05-25	Implantação concluída	Amazon EMR 5.36.1 totalmente implantado em todas as regiões suportadas
2023-05-09	Publicação de documentos	Notas de lançamento do Amazon EMR 5.36.1 publicadas pela primeira vez
2023-05-04	Lançamento inicial	Amazon EMR 5.36.1 foi implantado pela primeira vez em regiões comerciais limitadas

Amazon EMR versão 5.36.0

Versões 5.36.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.36.0	emr-5.35.0	emr-5.34.0	emr-5.33.1
AWS SDK para Java	1.12.206	1.12.159	1.11.970	1.11.970
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.11.12	2.11.12	2.11.12	2.11.12
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

Notas da versão 5.36.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.36.0 da Amazon. As alterações são referentes à versão 5.35.0.

Data da versão inicial: 15 de junho de 2022

Novos atributos

- A EMR versão 5.36.0 da Amazon adiciona suporte à linguagem de definição de dados (DDL) com o Apache Spark em clusters habilitados para Apache Ranger. Isso permite que você use o Apache Ranger para gerenciar o acesso para operações como criar, alterar e descartar bancos de dados e tabelas de um cluster da Amazon. EMR
- Quando você inicia um cluster com a versão de patch mais recente do Amazon EMR 5.36 ou superior, 6.6 ou superior ou 7.0 ou superior, a Amazon EMR usa a versão mais recente do Amazon Linux 2023 ou Amazon Linux 2 para a Amazon padrão. EMR AMI Para obter mais informações, consulte [Usando o Amazon Linux padrão AMI para a Amazon EMR](#).

Note

Esta versão não recebe mais AMI atualizações automáticas, pois foi sucedida por mais 1 lançamento de patch. A versão de patch é indicada pelo número após o segundo ponto decimal (6.8.**1**). Para ver se você está usando a versão de patch mais recente, verifique as versões disponíveis no [Guia de lançamento](#) ou verifique o menu suspenso de EMRlançamentos da Amazon ao criar um cluster no console ou usar a ação [ListReleaseLabels](#)API ou [list-release-labels](#)CLI. Para receber atualizações sobre novos lançamentos, assine o RSS feed do [What's new?](#) página.

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 709.1	4.14.348	23 de julho de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), AWS GovCloud China (Pequim), China AWS GovCloud (Ningxia), Ásia Pacífico (Hyderabad), Oriente Médio (), Europa (EspanhaUAE), Europa (Zurique), Ásia-Pacífico

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(Melbourne), Israel (Tel Aviv)
2.0.2023 504.1	4.14.313	16 de maio de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 418.0	4.14.311	3 de maio de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 404.1	4.14.311	18 de abril de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)
2.0.2023 404.0	4.14.311	10 de abril de 2023	Leste dos EUA (Norte da Virgínia), Europa (Paris)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 320.0	4.14.309	30 de março de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 307.0	4.14.305	15 de março de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 207.0	4.14.304	22 de fevereiro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 119.1	4.14.301	3 de fevereiro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 210.1	4.14.301	22 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 103.3	4.14.296	5 de dezembro de 2022	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 004.0	4.14.294	2 de novembro de 2022	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 912.1	4.14.291	7 de outubro de 2022	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)
2.0.2022 719.0	4.14.287	10 de agosto de 2022	Oeste dos EUA (Norte da Califórnia), Europa (Paris), Europa (Estocolmo), Europa (Frankfurt), Ásia-Pacífico (Mumbai), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 426.0	4.14.281	14 de junho de 2022	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

Alterações, melhorias e problemas resolvidos

- As atualizações do Amazon EMR 5.36.0 agora oferecem suporte a: 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, conector Amazon Glue 1.18.0, 2.51.0. EMRFS

Problemas conhecidos

- Quando você usa o Spark com a formatação de localização de partições do Hive para ler dados no Amazon S3 e executa o Spark nas EMR versões 5.30.0 a 5.36.0 e 6.2.0 a 6.9.0 da Amazon, você pode encontrar um problema que impede seu cluster de ler os dados corretamente. Isso poderá acontecer se suas partições tiverem todas as características a seguir:
 - Duas ou mais partições são verificadas na mesma tabela.
 - Pelo menos um caminho de diretório de partição é um prefixo de pelo menos outro caminho de diretório de partição, por exemplo, `s3://bucket/table/p=a` é um prefixo de `s3://bucket/table/p=a b`.
 - O primeiro caractere que segue o prefixo no outro diretório de partição tem um valor UTF -8 menor que o / caractere (U+002F). Por exemplo, o caractere de espaço (U+0020) que ocorre entre a e b em `s3://bucket/table/p=a b` se enquadra nessa categoria. Observe que existem 14 outros caracteres que não são de controle: `!"#$%&'()*+,-.` Para obter mais informações, consulte a [tabela de codificação UTF -8 e caracteres Unicode](#).

Como solução alternativa para esse problema, defina a configuração

`spark.sql.sources.fastS3PartitionDiscovery.enabled` como `false` na classificação `spark-defaults`.

- Com as EMR versões 5.36.0 e 6.6.0 a 6.9.0 da Amazon, `SecretAgent` os componentes do `RecordServer` serviço podem sofrer perda de dados de log devido a uma configuração incorreta do padrão de nome de arquivo nas propriedades do `Log4j2`. A configuração incorreta faz com que os componentes gerem somente um arquivo de log por dia. Quando a estratégia de rotação ocorre, ela substitui o arquivo existente em vez de gerar um novo arquivo de log, conforme esperado. Como solução alternativa, use uma ação de bootstrap para gerar arquivos de log a cada hora e acrescentar um número inteiro de incremento automático no nome do arquivo para lidar com a rotação.

Para as versões EMR 6.6.0 a 6.9.0 da Amazon, use a seguinte ação de bootstrap ao iniciar um cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Para o Amazon EMR 5.36.0, use a seguinte ação de bootstrap ao iniciar um cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

Versões de componente 5.36.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion-amzn-EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.16.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.16.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.5.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-notebook-env</code>	1.5.0	Ambiente Conda para bloco de anotações do emr, que

Componente	Version (Versão)	Descrição
		inclui o jupyter enterprise gateway
emr-s3-dist-cp	2.21.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	1.7.0	EMRConector S3Select
emrfs	2.51.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.14.2	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
flink-jobmanager-config	1.14.2	Gerenciando recursos em EMR nós para o Apache JobManager Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.

Componente	Version (Versão)	Descrição
hadoop-client	2.10.1-amzn-4	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.10.1-amzn-4	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.10.1-amzn-4	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.10.1-amzn-4	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	2.10.1-amzn-4	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httpfs-server	2.10.1-amzn-4	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.10.1-amzn-4	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.10.1-amzn-4	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.10.1-amzn-4	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.10.1-amzn-4	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.

Componente	Version (Versão)	Descrição
hadoop-yarn-timeline-server	2.10.1-amzn-4	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.13	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.13	Serviço para atender uma ou mais HBase regiões.
hbase-client	1.4.13	HBase cliente de linha de comando.
hbase-rest-server	1.4.13	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.4.13	Serviço que fornece um endpoint Thrift para HBase.
hcatalog-client	2,3.9-amzn-2	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2,3.9-amzn-2	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2,3.9-amzn-2	HTTP endpoint fornecendo uma REST interface para HCatalog.

Componente	Version (Versão)	Descrição
hive-client	2,3.9-amzn-2	O cliente da linha de comando do Hive.
hive-hbase	2,3.9-amzn-2	Cliente do Hive-hbase.
hive-metastore-server	2,3.9-amzn-2	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2,3.9-amzn-2	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.10.1-amzn-1	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.
hudi-spark	0.10.1-amzn-1	Biblioteca de pacotes para executar o Spark com o Hudi.
hudi-presto	0.10.1-amzn-1	Biblioteca de pacotes para executar o Presto com o Hudi.
hue-server	4.10.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	1.4.1	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.7.1-incubating	RESTinterface para interagir com o Apache Spark

Componente	Version (Versão)	Descrição
nginx	1.13.1	nginx [engine x] é HTTP um servidor proxy reverso
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.8.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mariadb-server	5.5.68	Meu servidor SQL de banco de dados.
nvidia-cuda	11.0.194	Drivers NVIDIA e toolkit CUDA
oozie-client	5.2.1	O cliente da linha de comando do Oozie.
oozie-server	5.2.1	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	4.5.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.3- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.3- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.267-amzn-1	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.

Componente	Version (Versão)	Descrição
presto-worker	0.267-amzn-1	O serviço que executa partes de uma consulta.
presto-client	0.267-amzn-1	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	4.0.2	O projeto R para computação estatística
ranger-kms-server	1.2.0	Sistema de gerenciamento de chaves Apache Ranger
spark-client	2.4.8-amzn-2	Os clientes da linha de comando do Spark.
spark-history-server	2.4.8-amzn-2	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.4.8-amzn-2	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.4.8-amzn-2	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.

Componente	Version (Versão)	Descrição
tensorflow	2.4.1	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.2	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.10.0	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.14	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.14	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.36.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Ocorrem ações de reconfiguração quando você especifica uma configuração para grupos de instâncias em um cluster em execução. A Amazon EMR só inicia ações de reconfiguração para as classificações que você modifica. Para ter mais informações, consulte [Reconfigurar um grupo de instâncias de um cluster em execução](#).

Classificações do emr-5.36.0

Classificações	Descrição	Ações de reconfiguração
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.	Restarts the ResourceManager service.
container-executor	Altere os valores no arquivo container-executor.cfg YARN do Hadoop.	Not available.
container-log4j	Altere os valores no arquivo container-log4j.properties YARN do Hadoop.	Not available.
core-site	Alterar os valores no arquivo core-site.xml do 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	Alterar as configurações relacionadas ao docker.	Not available.
emrfs-site	Altere EMRFS as configurações.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

Classificações	Descrição	Ações de reconfiguração
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Alterar as configurações do flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Alterar as configurações de log4j.properties no Flink.	Restarts Flink history server.
flink-log4j-session	Alterar as configurações de log4j-yarn-session.properties no Flink para sessão do Kubernetes/Yarn.	Not available.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.	Restarts Flink history server.

Classificações	Descrição	Ações de reconfiguração
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do 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	Alterar os valores no arquivo log4j.properties do 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	Alterar a configuração do servidor SSL no Hadoop	Not available.
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop	Not available.

Classificações	Descrição	Ações de reconfiguração
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Mude os valores no HBase ambiente.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.	Not available.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.	Should not be reconfigured.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Classificações	Descrição	Ações de reconfiguração
hcatalog-env	Mude os valores no HCatalog ambiente.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.	Not available.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.	Not available.
hive-env	Alterar os valores no ambiente do Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log 4j2.properties do Hive.	Restarts HiveServer2 and HiveMetastore.

Classificações	Descrição	Ações de reconfiguração
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log 4j2.properties do Hive.	Not available.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.	Not available.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.	Not available.
hue-ini	Alterar os valores no arquivo ini do Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Mude os valores no HTTPFS ambiente.	Restarts Hadoop Httpfs service.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.	Not available.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.	Not available.

Classificações	Descrição	Ações de reconfiguração
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Altere os valores no ambiente do Hudi.	Not available.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.	Not available.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.	Not available.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.	Not available.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.	Not available.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.	Restarts Livy Server.
livy-env	Alterar os valores no ambiente do Livy.	Restarts Livy Server.
livy-log4j	Alterar as configurações de log4j.properties no Livy.	Restarts Livy Server.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.

Classificações	Descrição	Ações de reconfiguração
oozie-env	Alterar os valores no ambiente do Oozie.	Restarts Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.	Restarts Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.	Not available.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.	Not available.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.	Not available.
pig-env	Alterar os valores no ambiente do Pig.	Not available.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.	Restarts Oozie.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.	Not available.
presto-log	Alterar os valores no arquivo log.properties do Presto.	Restarts Presto-Server.

Classificações	Descrição	Ações de reconfiguração
presto-config	Alterar os valores no arquivo <code>config.properties</code> do Presto.	Restarts Presto-Server.
presto-password-authenticator	Altere os valores no arquivo <code>password-authenticator.properties</code> do Presto.	Not available.
presto-env	Altere os valores no arquivo <code>presto-env.sh</code> do Presto.	Restarts Presto-Server.
presto-node	Altere os valores no arquivo <code>node.properties</code> do Presto.	Not available.
presto-connector-blackhole	Alterar os valores no arquivo <code>blackhole.properties</code> do Presto.	Not available.
presto-connector-cassandra	Alterar os valores no arquivo <code>cassandra.properties</code> do Presto.	Not available.
presto-connector-hive	Alterar os valores no arquivo <code>hive.properties</code> do Presto.	Restarts Presto-Server.
presto-connector-jmx	Alterar os valores no arquivo <code>jmx.properties</code> do Presto.	Not available.
presto-connector-kafka	Alterar os valores no arquivo <code>kafka.properties</code> do Presto.	Not available.
presto-connector-localfile	Alterar os valores no arquivo <code>localfile.properties</code> do Presto.	Not available.
presto-connector-memory	Alterar os valores no arquivo <code>memory.properties</code> do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.	Not available.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.	Not available.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.	Not available.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.	Not available.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.	Not available.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.	Not available.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.	Not available.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.	Not available.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.	Restarts Ranger KMS Server.

Classificações	Descrição	Ações de reconfiguração
ranger-kms-log4j	Altere os valores no arquivo kms-log4j.properties do Ranger. KMS	Not available.
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.	Not available.
recordserver-env	Mude os valores no EMR RecordServer ambiente.	Restarts EMR record server.
recordserver-conf	Altere os valores no arquivo EMR RecordServer erver.properties do.	Restarts EMR record server.
recordserver-log4j	Altere os valores no EMR RecordServer arquivo log4j.properties.	Restarts EMR record server.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Alterar os valores no ambiente do Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.	Not available.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.	Restarts Spark history server and Spark thrift server.

Classificações	Descrição	Ações de reconfiguração
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Alterar os valores no ambiente do Sqoop.	Not available.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.	Not available.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.	Not available.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.	Restarts Oozie.
yarn-env	Mude os valores no YARN ambiente.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.	Restarts Zeppelin.
zeppelin-site	Alterar as definições da configuração em zeppelin-site.xml.	Restarts Zeppelin.

Classificações	Descrição	Ações de reconfiguração
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.	Restarts Zookeeper server.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.	Restarts Zookeeper server.

Amazon EMR versão 5.35.0

Versões 5.35.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [Iceberg](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.35.0	emr-5.34.0	emr-5.33.1	emr-5.33.0
AWS SDK para Java	1.12.159	1.11.970	1.11.970	1.11.970
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.11.12	2.11.12	2.11.12	2.11.12

	emr-5.35.0	emr-5.34.0	emr-5.33.1	emr-5.33.0
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
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

	emr-5.35.0	emr-5.34.0	emr-5.33.1	emr-5.33.0
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
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

Notas da versão 5.35.0

Esta é a nota de EMR lançamento da Amazon versão 5.35.0.

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.35.0 da Amazon. As alterações são referentes à versão 5.34.0.

Data da versão inicial: 30 de março de 2022

Novos atributos

- Os aplicativos da Amazon EMR versão 5.35 que usam o Log4j 1.x e o Log4j 2.x são atualizados para usar o Log4j 1.2.17 (ou superior) e o Log4j 2.17.1 (ou superior), respectivamente, e não exigem o uso de ações de bootstrap para mitigar os problemas nas versões anteriores. CVE Consulte [Abordagem para CVE mitigar -2021-44228](#).

Alterações, melhorias e problemas resolvidos

Alterações no Flink

Alterar tipo	Descrição
Atualizações	<ul style="list-style-type: none"> • Atualizar a versão do flink para 1.14.2. • log4j atualizado para 2.17.1.

Alterações no Hadoop

Alterar tipo	Descrição
Backports de código aberto do Hadoop desde 5.34.0 EMR	<ul style="list-style-type: none"> • YARN-10438: Manipulação de nulo container Id em C # () lientRMService getContainerReport • YARN-7266: Tópicos do manipulador de eventos do Timeline Server bloqueados • YARN-10438: ATS 1.5 falha ao iniciar se os RollingLevelDb arquivos estiverem corrompidos ou ausentes • HADOOP-13500: Sincronizando a iteração do objeto de propriedades de configuração • YARN-10651: CapacityScheduler caiu com o in. NPE AbstractYarnScheduler updateNodeResource() • HDFS-12221: Substituir xerces em XmlEditsVisitor • HDFS-16410: Análise insegura de Xml em OfflineEditsXMLLoader
Alterações e correções no Hadoop	<ul style="list-style-type: none"> • O Tomcat usado em KMS e o HTTPFs foi atualizado para 8.5.75 • Na FileSystemOptimizedCommitter V2, o marcador de sucesso foi escrito no caminho de commitJob saída definido durante a criação do committer. Como commitJob os caminhos de saída no nível da tarefa

Alterar tipo	Descrição
	<p>podem ser diferentes, o caminho foi corrigido para usar aquele definido nos arquivos de manifesto. Para trabalhos do Hive, isso faz com que o marcador de sucesso seja gravado corretamente ao executar operações como partição dinâmica ou UNION ALL</p>

Alterações no Hive

Alterar tipo	Descrição
<p>O Hive foi atualizado para a versão 2.3.9 de código aberto, incluindo essas correções JIRA</p>	<ul style="list-style-type: none"> • HIVE-17155: findConfFile () em HiveConf .java tem alguns problemas com o caminho de configuração • HIVE-24797: Desative a validação de valores padrão ao analisar esquemas Avro • HIVE-21563: Melhore o desempenho do getEmptyTable Table# desativando Once registerAllFunctions • HIVE-18147: Os testes podem falhar com java.net. BindException: Endereço já em uso • HIVE-24608: Volte para get_table no cliente para Hive 2.3.x HMS • HIVE-21200: Vetorização - coluna de data lançando java.lang. UnsupportedOperationException para parquet • HIVE-19228: Remover o uso do commons-httpclient 3.x
<p>Backports de código aberto do Hive desde 5.34.0 EMR</p>	<ul style="list-style-type: none"> • HIVE-19990: Falha na consulta com intervalo literal na condição de junção • HIVE-25824: Atualize o branch-2.3 para o log4j 2.17.0

Alterar tipo	Descrição
	<ul style="list-style-type: none"> • TEZ-4062: O agendamento de tentativas especulativas deve ser abortado quando a tarefa for concluída • TEZ-4108: NullPointerException durante a execução especulativa, condição de corrida • TEZ-3918: A configuração de <code>tez.task.log.level</code> não funciona
Atualizações e correções do Hive	<ul style="list-style-type: none"> • Atualizar a versão Log4j para 2.17.1 • Atualize a ORC versão para 1.4.3 • Corrigido o impasse devido à discussão de penalidades em ShuffleScheduler
Novos atributos	<ul style="list-style-type: none"> • Adicionado atributo para imprimir consultas do Hive em logs AM. Ela fica desabilitada por padrão. Sinalizador/Conf.: <code>tez.am.emr.print.hive.query.in.log</code> . Status (padrão):FALSE.

Alterações no Oozie

Alterar tipo	Descrição
Backports de código aberto Oozie desde 5.34.0 EMR	<ul style="list-style-type: none"> • OOZIE-3652: O lançador Oozie deve tentar novamente a listagem de diretórios quando ocorrer <code>NoSuchFileException</code>

Alterações no Pig

Alterar tipo	Descrição
Atualizações	<ul style="list-style-type: none"> • log4j atualizado para 1.2.17.

Problemas conhecidos

- Quando você usa o Spark com a formatação de localização de partições do Hive para ler dados no Amazon S3 e executa o Spark nas EMR versões 5.30.0 a 5.36.0 e 6.2.0 a 6.9.0 da Amazon, você pode encontrar um problema que impede seu cluster de ler os dados corretamente. Isso poderá acontecer se suas partições tiverem todas as características a seguir:
 - Duas ou mais partições são verificadas na mesma tabela.
 - Pelo menos um caminho de diretório de partição é um prefixo de pelo menos outro caminho de diretório de partição, por exemplo, `s3://bucket/table/p=a` é um prefixo de `s3://bucket/table/p=a b`.
 - O primeiro caractere que segue o prefixo no outro diretório de partição tem um valor UTF -8 menor que o `/` caractere (U+002F). Por exemplo, o caractere de espaço (U+0020) que ocorre entre `a` e `b` em `s3://bucket/table/p=a b` se enquadra nessa categoria. Observe que existem 14 outros caracteres que não são de controle: `!"#$%&'()*+,-.` Para obter mais informações, consulte a [tabela de codificação UTF -8 e caracteres Unicode](#).

Como solução alternativa para esse problema, defina a configuração `spark.sql.sources.fastS3PartitionDiscovery.enabled` como `false` na classificação `spark-defaults`.

Versões de componente 5.35.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK
emr-ddb	4.16.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.15.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.5.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-notebook-env	1.5.0	Ambiente Conda para bloco de anotações do emr, que inclui o jupyter enterprise gateway
emr-s3-dist-cp	2.20.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	1.7.0	EMRConector S3Select
emrfs	2.49.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.14.2	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
flink-jobmanager-config	1.14.2	Gerenciando recursos em EMR nós para o Apache JobManager Flink.

Componente	Version (Versão)	Descrição
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.10.1-amzn-3	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.10.1-amzn-3	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.10.1-amzn-3	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.10.1-amzn-3	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	2.10.1-amzn-3	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httpfs-server	2.10.1-amzn-3	HTTPendpoint para HDFS operações.

Componente	Version (Versão)	Descrição
hadoop-kms-server	2.10.1-amzn-3	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.10.1-amzn-3	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.10.1-amzn-3	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.10.1-amzn-3	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.10.1-amzn-3	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.13	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.13	Serviço para atender uma ou mais HBase regiões.
hbase-client	1.4.13	HBasecliente de linha de comando.
hbase-rest-server	1.4.13	Serviço que fornece um RESTful HTTP endpoint paraHBase.

Componente	Version (Versão)	Descrição
hbase-thrift-server	1.4.13	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	2.3.9-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.9-amzn-0	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.9-amzn-0	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	2.3.9-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.3.9-amzn-0	Cliente do Hive-hbase.
hive-metastore-server	2.3.9-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.9-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.9.0-amzn-2	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.

Componente	Version (Versão)	Descrição
hudi-spark	0.9.0-amzn-2	Biblioteca de pacotes para executar o Spark com o Hudi.
hudi-presto	0.9.0-amzn-2	Biblioteca de pacotes para executar o Presto com o Hudi.
hue-server	4.10.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	1.4.1	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.7.1-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.13.1	nginx [engine x] é HTTP um servidor proxy reverso
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.8.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mariadb-server	5.5.68	Meu servidor SQL de banco de dados.
nvidia-cuda	10.1.243	Drivers NVIDIA e toolkit CUDA
oozie-client	5.2.1	O cliente da linha de comando do Oozie.

Componente	Version (Versão)	Descrição
oozie-server	5.2.1	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	4.5.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.3- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.3- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.266-amzn-0	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.266-amzn-0	O serviço que executa partes de uma consulta.
presto-client	0.266-amzn-0	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	4.0.2	O projeto R para computação estatística

Componente	Version (Versão)	Descrição
ranger-kms-server	1.2.0	Sistema de gerenciamento de chaves Apache Ranger
spark-client	2.4.8-amzn-1	Os clientes da linha de comando do Spark.
spark-history-server	2.4.8-amzn-1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.4.8-amzn-1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.4.8-amzn-1	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	2.4.1	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.2	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.
zeppelin-server	0.10.0	O notebook baseado na web que permite a análise de dados interativa.

Componente	Version (Versão)	Descrição
zookeeper-server	3.4.14	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.14	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.35.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Ocorrem ações de reconfiguração quando você especifica uma configuração para grupos de instâncias em um cluster em execução. A Amazon EMR só inicia ações de reconfiguração para as classificações que você modifica. Para ter mais informações, consulte [Reconfigurar um grupo de instâncias de um cluster em execução](#).

Classificações do emr-5.35.0

Classificações	Descrição	Ações de reconfiguração
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.	Restarts the Resource Manager service.
container-executor	Altere os valores no arquivo <code>container-executor.cfg</code> YARN do Hadoop.	Not available.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.	Not available.

Classificações	Descrição	Ações de reconfiguração
core-site	Alterar os valores no arquivo core-site.xml do 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	Alterar as configurações relacionadas ao docker.	Not available.
emrfs-site	Altere EMRFS as configurações.	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.

Classificações	Descrição	Ações de reconfiguração
flink-conf	Alterar as configurações do flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Alterar as configurações de log4j.properties no Flink.	Restarts Flink history server.
flink-log4j-session	Alterar as configurações de log4j-yarn-session.properties no Flink para sessão do Kubernetes/Yarn.	Not available.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.	Restarts Flink history server.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do 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.

Classificações	Descrição	Ações de reconfiguração
hadoop-log4j	Alterar os valores no arquivo log4j.properties do 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	Alterar a configuração do servidor SSL no Hadoop	Not available.
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop	Not available.
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Mude os valores no HBase ambiente.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.

Classificações	Descrição	Ações de reconfiguração
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.	Not available.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.	Should not be reconfigured.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Mude os valores no HCatalog ambiente.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat Server.

Classificações	Descrição	Ações de reconfiguração
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.	Not available.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.	Not available.
hive-env	Alterar os valores no ambiente do Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log 4j2.properties do Hive.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log 4j2.properties do Hive.	Not available.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.	Not available.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.	Not available.

Classificações	Descrição	Ações de reconfiguração
hue-ini	Alterar os valores no arquivo ini do Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Mude os valores no HTTPFS ambiente.	Restarts Hadoop Httpfs service.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.	Not available.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.	Not available.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Altere os valores no ambiente do Hudi.	Not available.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.	Not available.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.	Not available.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.	Not available.

Classificações	Descrição	Ações de reconfiguração
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.	Not available.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.	Restarts Livy Server.
livy-env	Alterar os valores no ambiente do Livy.	Restarts Livy Server.
livy-log4j	Alterar as configurações de log4j.properties no Livy.	Restarts Livy Server.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Alterar os valores no ambiente do Oozie.	Restarts Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.	Restarts Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.	Not available.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.	Not available.

Classificações	Descrição	Ações de reconfiguração
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.	Not available.
pig-env	Alterar os valores no ambiente do Pig.	Not available.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.	Restarts Oozie.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.	Not available.
presto-log	Alterar os valores no arquivo log.properties do Presto.	Restarts Presto-Server.
presto-config	Alterar os valores no arquivo config.properties do Presto.	Restarts Presto-Server.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.	Not available.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.	Restarts Presto-Server.
presto-node	Altere os valores no arquivo node.properties do Presto.	Not available.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-cassandra	Alterar os valores no arquivo <code>cassandra.properties</code> do Presto.	Not available.
presto-connector-hive	Alterar os valores no arquivo <code>hive.properties</code> do Presto.	Restarts Presto-Server.
presto-connector-jmx	Alterar os valores no arquivo <code>jmx.properties</code> do Presto.	Not available.
presto-connector-kafka	Alterar os valores no arquivo <code>kafka.properties</code> do Presto.	Not available.
presto-connector-localfile	Alterar os valores no arquivo <code>localfile.properties</code> do Presto.	Not available.
presto-connector-memory	Alterar os valores no arquivo <code>memory.properties</code> do Presto.	Not available.
presto-connector-mongodb	Alterar os valores no arquivo <code>mongodb.properties</code> do Presto.	Not available.
presto-connector-mysql	Alterar os valores no arquivo <code>mysql.properties</code> do Presto.	Not available.
presto-connector-postgresql	Alterar os valores no arquivo <code>postgresql.properties</code> do Presto.	Not available.
presto-connector-raptor	Alterar os valores no arquivo <code>raptor.properties</code> do Presto.	Not available.
presto-connector-redis	Alterar os valores no arquivo <code>redis.properties</code> do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.	Not available.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.	Not available.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.	Not available.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Altere os valores no arquivo kms-log4j.properties do Ranger. KMS	Not available.
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.	Not available.
recordserver-env	Mude os valores no EMR RecordServer ambiente.	Restarts EMR record server.
recordserver-conf	Altere os valores no arquivo EMR RecordServer erver.properties do.	Restarts EMR record server.

Classificações	Descrição	Ações de reconfiguração
recordserver-log4j	Altere os valores no EMR RecordServer arquivo log4j.properties.	Restarts EMR record server.
spark	Configurações EMR seleciona das pela Amazon para o Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Alterar os valores no ambiente do Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.	Not available.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Alterar os valores no ambiente do Sqoop.	Not available.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.	Not available.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.	Not available.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.	Restarts Oozie.

Classificações	Descrição	Ações de reconfiguração
yarn-env	Mude os valores no YARN ambiente.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.	Restarts Zeppelin.
zeppelin-site	Alterar as definições da configuração em zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.	Restarts Zookeeper server.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.	Restarts Zookeeper server.

Amazon EMR versão 5.34.0

Versões 5.34.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.34.0	emr-5.33.1	emr-5.33.0	emr-5.32.1
AWS SDK para Java	1.11.970	1.11.970	1.11.970	1.11.890
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.11.12	2.11.12	2.11.12	2.11.12
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
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

	emr-5.34.0	emr-5.33.1	emr-5.33.0	emr-5.32.1
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
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
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

Notas da versão 5.34.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.34.0 da Amazon. As alterações são referentes à versão 5.33.1.

Data da versão inicial: 20 de janeiro de 2022

Data da versão atualizada: 21 de março de 2022

Novos atributos

- [Escalabilidade gerenciada] Otimização de escalabilidade gerenciada de dados do Spark shuffle - Para EMR as versões 5.34.0 e posteriores da Amazon e as EMR versões 6.4.0 e posteriores, a escalabilidade gerenciada agora reconhece os dados do Spark shuffle (dados que o Spark redistribui entre partições para realizar operações específicas). Para obter mais informações sobre operações aleatórias, consulte [Usando escalabilidade EMR gerenciada na Amazon no Guia de EMR Gerenciamento EMR da Amazon](#) e no Guia de Programação do [Spark](#).
- [Hudi] Melhorias para simplificar a configuração do Hudi. Desabilitado o controle de simultaneidade otimista por padrão.

Alterações, melhorias e problemas resolvidos

- Esta é uma versão para corrigir problemas com o Amazon EMR Scaling quando ele falha na escalabilidade ou redução de escala de um cluster com sucesso ou causa falhas no aplicativo.
- Anteriormente, a reinicialização manual do gerenciador de recursos em um cluster multimestre fazia com que os daemons EMR no cluster da Amazon, como o Zookeeper, recarregassem todos os nós anteriormente desativados ou perdidos no arquivo znode do Zookeeper. Isso fez com que os limites padrão fossem excedidos em determinadas situações. A Amazon EMR agora remove os registros de nós desativados ou perdidos há mais de uma hora do arquivo do Zookeeper e os limites internos foram aumentados.
- Corrigido um problema em que as solicitações de escalabilidade falhavam em um cluster grande e altamente utilizado quando os daemons no cluster da EMR Amazon estavam executando atividades de verificação de integridade, como coletar o estado do nó e o estado do YARN nó. HDFS Isso estava acontecendo porque os daemons no cluster não conseguiam comunicar os dados do status de saúde de um nó aos componentes internos da Amazon. EMR
- Daemons EMR no cluster aprimorados para rastrear corretamente os estados dos nós quando os endereços IP são reutilizados para melhorar a confiabilidade durante as operações de escalabilidade.

- [SPARK-29683](#). Corrigido um problema em que ocorriam falhas de trabalho durante a redução da escala verticalmente do cluster, pois o Spark presumia que todos os nós disponíveis estavam na lista de negação.
- [YARN-9011](#). Corrigido um problema em que falhas de trabalho ocorriam devido a uma condição de corrida no YARN descomissionamento quando o cluster tentava aumentar ou diminuir a escala.
- Foi corrigido o problema com falhas de etapas ou tarefas durante a escalabilidade do cluster, garantindo que os estados dos nós fossem sempre consistentes entre os EMR daemons no cluster da Amazon e/. YARN HDFS
- Corrigido um problema em que operações de cluster, como redução de escala e envio de etapas, falhavam em EMR clusters da Amazon habilitados com a autenticação Kerberos. Isso ocorreu porque o daemon no EMR cluster da Amazon não renovou o tíquete Kerberos, que é necessário para se comunicar de forma segura com/em execução no nó primário. HDFS YARN
- Zeppelin atualizado para a versão 0.10.0.
- Livy Fix: atualizado para 0.7.1
- Melhoria do desempenho do Spark - executores heterogêneos são desativados quando determinados valores de configuração do Spark são substituídos na versão 5.34.0. EMR
- Os servidores Web HDFS e HTTPFs estão desativados por padrão. Você pode reativar a Web HDFS usando a configuração do Hadoop, `dfs.webhdfs.enabled` O servidor HttpFS pode ser iniciado usando `sudo systemctl start hadoop-httpfs`.

Problemas conhecidos

- O recurso Amazon EMR Notebooks usado com a representação de usuário do Livy não funciona porque o HTTPFs está desativado por padrão. Nesse caso, o EMR notebook não pode se conectar ao cluster que tem a representação Livy ativada. A solução alternativa é iniciar o servidor HTTPFs antes de conectar o EMR notebook ao cluster usando o `sudo systemctl start hadoop-httpfs`
- As consultas Hue não funcionam no Amazon EMR 6.4.0 porque o servidor Apache Hadoop HTTPFs está desativado por padrão. Para usar o Hue na Amazon EMR 6.4.0, inicie manualmente o servidor HTTPFs no nó EMR primário da Amazon usando `sudo systemctl start hadoop-httpfs` ou use [uma etapa da Amazon](#). EMR
- O recurso Amazon EMR Notebooks usado com a representação de usuário do Livy não funciona porque o HTTPFs está desativado por padrão. Nesse caso, o EMR notebook não pode se conectar ao cluster que tem a representação Livy ativada. A solução alternativa é iniciar o servidor HTTPFs

antes de conectar o EMR notebook ao cluster usando o. `sudo systemctl start hadoop-
httpfs`

- Quando você usa o Spark com a formatação de localização de partições do Hive para ler dados no Amazon S3 e executa o Spark nas EMR versões 5.30.0 a 5.36.0 e 6.2.0 a 6.9.0 da Amazon, você pode encontrar um problema que impede seu cluster de ler os dados corretamente. Isso poderá acontecer se suas partições tiverem todas as características a seguir:
 - Duas ou mais partições são verificadas na mesma tabela.
 - Pelo menos um caminho de diretório de partição é um prefixo de pelo menos outro caminho de diretório de partição, por exemplo, `s3://bucket/table/p=a` é um prefixo de `s3://bucket/table/p=a b`.
 - O primeiro caractere que segue o prefixo no outro diretório de partição tem um valor UTF -8 menor que o / caractere (U+002F). Por exemplo, o caractere de espaço (U+0020) que ocorre entre a e b em `s3://bucket/table/p=a b` se enquadra nessa categoria. Observe que existem 14 outros caracteres que não são de controle: `!"#$%&'()*+,-.` Para obter mais informações, consulte a [tabela de codificação UTF -8 e caracteres Unicode](#).

Como solução alternativa para esse problema, defina a configuração `spark.sql.sources.fastS3PartitionDiscovery.enabled` como `false` na classificação `spark-defaults`.

Versões de componente 5.34.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK
emr-ddb	4.16.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.14.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.5.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-notebook-env	1.4.0	Ambiente Conda para bloco de anotações do emr, que inclui o jupyter enterprise gateway
emr-s3-dist-cp	2.18.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	1.7.0	EMRConector S3Select
emrfs	2.48.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.13.1	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
flink-jobmanager-config	1.13.1	Gerenciando recursos em EMR nós para o Apache JobManager Flink.

Componente	Version (Versão)	Descrição
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.10.1-amzn-2	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.10.1-amzn-2	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.10.1-amzn-2	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.10.1-amzn-2	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	2.10.1-amzn-2	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httpfs-server	2.10.1-amzn-2	HTTPendpoint para HDFS operações.

Componente	Version (Versão)	Descrição
hadoop-kms-server	2.10.1-amzn-2	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.10.1-amzn-2	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.10.1-amzn-2	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.10.1-amzn-2	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.10.1-amzn-2	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.13	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.13	Serviço para atender uma ou mais HBase regiões.
hbase-client	1.4.13	HBasecliente de linha de comando.
hbase-rest-server	1.4.13	Serviço que fornece um RESTful HTTP endpoint paraHBase.

Componente	Version (Versão)	Descrição
hbase-thrift-server	1.4.13	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	2.3.8-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.8-amzn-0	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.8-amzn-0	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	2.3.8-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.3.8-amzn-0	Cliente do Hive-hbase.
hive-metastore-server	2.3.8-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.8-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.9.0-amzn-0	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.

Componente	Version (Versão)	Descrição
hudi-spark	0.9.0-amzn-0	Biblioteca de pacotes para executar o Spark com o Hudi.
hudi-presto	0.9.0-amzn-0	Biblioteca de pacotes para executar o Presto com o Hudi.
hue-server	4.9.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	1.4.1	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.7.1-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.13.1	nginx [engine x] é HTTP um servidor proxy reverso
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.8.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mariadb-server	5.5.68	Meu servidor SQL de banco de dados.
nvidia-cuda	10.1.243	Drivers NVIDIA e toolkit CUDA
oozie-client	5.2.1	O cliente da linha de comando do Oozie.

Componente	Version (Versão)	Descrição
oozie-server	5.2.1	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	4.5.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.3- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.3- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.261-amzn-0	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.261-amzn-0	O serviço que executa partes de uma consulta.
presto-client	0.261-amzn-0	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	4.0.2	O projeto R para computação estatística

Componente	Version (Versão)	Descrição
ranger-kms-server	1.2.0	Sistema de gerenciamento de chaves Apache Ranger
spark-client	2.4.8-amzn-0	Os clientes da linha de comando do Spark.
spark-history-server	2.4.8-amzn-0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.4.8-amzn-0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.4.8-amzn-0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	2.4.1	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.2	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.
zeppelin-server	0.10.0	O notebook baseado na web que permite a análise de dados interativa.

Componente	Version (Versão)	Descrição
zookeeper-server	3.4.14	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.14	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.34.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Ocorrem ações de reconfiguração quando você especifica uma configuração para grupos de instâncias em um cluster em execução. A Amazon EMR só inicia ações de reconfiguração para as classificações que você modifica. Para ter mais informações, consulte [Reconfigurar um grupo de instâncias de um cluster em execução](#).

Classificações do emr-5.34.0

Classificações	Descrição	Ações de reconfiguração
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.	Restarts the Resource Manager service.
container-executor	Altere os valores no arquivo <code>container-executor.cfg</code> YARN do Hadoop.	Not available.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.	Not available.

Classificações	Descrição	Ações de reconfiguração
core-site	Alterar os valores no arquivo core-site.xml do 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	Alterar as configurações relacionadas ao docker.	Not available.
emrfs-site	Altere EMRFS as configurações.	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.

Classificações	Descrição	Ações de reconfiguração
flink-conf	Alterar as configurações do flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Alterar as configurações de log4j.properties no Flink.	Restarts Flink history server.
flink-log4j-session	Alterar as configurações de log4j-yarn-session.properties no Flink para sessão do Kubernetes/Yarn.	Not available.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.	Restarts Flink history server.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do 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.

Classificações	Descrição	Ações de reconfiguração
hadoop-log4j	Alterar os valores no arquivo log4j.properties do 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	Alterar a configuração do servidor SSL no Hadoop	Not available.
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop	Not available.
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Mude os valores no HBase ambiente.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.

Classificações	Descrição	Ações de reconfiguração
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.	Not available.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.	Should not be reconfigured.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Mude os valores no HCatalog ambiente.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat Server.

Classificações	Descrição	Ações de reconfiguração
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.	Not available.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.	Not available.
hive-env	Alterar os valores no ambiente do Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log 4j2.properties do Hive.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log 4j2.properties do Hive.	Not available.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.	Not available.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.	Not available.

Classificações	Descrição	Ações de reconfiguração
hue-ini	Alterar os valores no arquivo ini do Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Mude os valores no HTTPFS ambiente.	Restarts Hadoop Httpfs service.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.	Not available.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.	Not available.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Altere os valores no ambiente do Hudi.	Not available.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.	Not available.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.	Not available.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.	Not available.

Classificações	Descrição	Ações de reconfiguração
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.	Not available.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.	Restarts Livy Server.
livy-env	Alterar os valores no ambiente do Livy.	Restarts Livy Server.
livy-log4j	Alterar as configurações de log4j.properties no Livy.	Restarts Livy Server.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Alterar os valores no ambiente do Oozie.	Restarts Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.	Restarts Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.	Not available.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.	Not available.

Classificações	Descrição	Ações de reconfiguração
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.	Not available.
pig-env	Alterar os valores no ambiente do Pig.	Not available.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.	Restarts Oozie.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.	Not available.
presto-log	Alterar os valores no arquivo log.properties do Presto.	Restarts Presto-Server.
presto-config	Alterar os valores no arquivo config.properties do Presto.	Restarts Presto-Server.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.	Not available.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.	Restarts Presto-Server.
presto-node	Altere os valores no arquivo node.properties do Presto.	Not available.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-cassandra	Alterar os valores no arquivo <code>cassandra.properties</code> do Presto.	Not available.
presto-connector-hive	Alterar os valores no arquivo <code>hive.properties</code> do Presto.	Restarts Presto-Server.
presto-connector-jmx	Alterar os valores no arquivo <code>jmx.properties</code> do Presto.	Not available.
presto-connector-kafka	Alterar os valores no arquivo <code>kafka.properties</code> do Presto.	Not available.
presto-connector-localfile	Alterar os valores no arquivo <code>localfile.properties</code> do Presto.	Not available.
presto-connector-memory	Alterar os valores no arquivo <code>memory.properties</code> do Presto.	Not available.
presto-connector-mongodb	Alterar os valores no arquivo <code>mongodb.properties</code> do Presto.	Not available.
presto-connector-mysql	Alterar os valores no arquivo <code>mysql.properties</code> do Presto.	Not available.
presto-connector-postgresql	Alterar os valores no arquivo <code>postgresql.properties</code> do Presto.	Not available.
presto-connector-raptor	Alterar os valores no arquivo <code>raptor.properties</code> do Presto.	Not available.
presto-connector-redis	Alterar os valores no arquivo <code>redis.properties</code> do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.	Not available.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.	Not available.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.	Not available.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Altere os valores no arquivo kms-log4j.properties do Ranger. KMS	Not available.
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.	Not available.
recordserver-env	Mude os valores no EMR RecordServer ambiente.	Restarts EMR record server.
recordserver-conf	Altere os valores no arquivo EMR RecordServer erver.properties do.	Restarts EMR record server.

Classificações	Descrição	Ações de reconfiguração
recordserver-log4j	Altere os valores no EMR RecordServer arquivo log4j.properties.	Restarts EMR record server.
spark	Configurações EMR seleciona das pela Amazon para o Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Alterar os valores no ambiente do Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.	Not available.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Alterar os valores no ambiente do Sqoop.	Not available.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.	Not available.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.	Not available.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.	Restarts Oozie.

Classificações	Descrição	Ações de reconfiguração
yarn-env	Mude os valores no YARN ambiente.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.	Restarts Zeppelin.
zeppelin-site	Alterar as definições da configuração em zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.	Restarts Zookeeper server.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.	Restarts Zookeeper server.

Amazon EMR versão 5.33.1

Versões 5.33.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.33.1	emr-5.33.0	emr-5.32.1	emr-5.32.0
AWS SDK para Java	1.11.970	1.11.970	1.11.890	1.11.890
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.11.12	2.11.12	2.11.12	2.11.12
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
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

	emr-5.33.1	emr-5.33.0	emr-5.32.1	emr-5.32.0
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
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

Notas da versão 5.33.1

As notas de lançamento a seguir incluem informações sobre a versão EMR 5.33.0/5.33.1 da Amazon. As alterações são referentes à versão 5.32.0.

Data da versão inicial: 19 de abril de 2021

Data da última atualização: 9 de agosto de 2021

Atualizações

- Atualizado o conector do Amazon Glue para a versão 1.15.0
- Atualizado para AWS SDK for Java a versão 1.11.970
- Atualizado para EMRFS a versão 2.46.0
- EMRGoodies atualizados para a versão 2.14.0
- Servidor de EMR registros atualizado para a versão 1.9.0
- EMRS3 Dist CP atualizado para a versão 2.18.0
- Agente EMR secreto atualizado para a versão 1.8.0
- Atualizado o Flink para a versão 1.12.1
- Atualizado o Hadoop para a versão 2.10.1-amzn-1
- Atualizado o Hive para a versão 2.3.7-amzn-4
- Atualizado o Hudi para a versão 0.7.0
- Atualizado o Hue para a versão 4.9.0
- Atualizado o OpenCV para a versão 4.5.0
- Atualizado o Presto para a versão 0.245.1-amzn-0
- Atualizado o R para a versão 4.0.2
- Atualizado o Spark para a versão 2.4.7-amzn-1
- Atualizado TensorFlow para a versão 2.4.1
- Atualizado o Zeppelin para a versão 0.9.0

Alterações, melhorias e problemas resolvidos

- Esta é uma versão para corrigir problemas com o Amazon EMR Scaling quando ele falha na escalabilidade ou redução de escala de um cluster com sucesso ou causa falhas no aplicativo.

- Corrigido um problema em que as solicitações de escalabilidade falhavam em um cluster grande e altamente utilizado quando os daemons no cluster da EMR Amazon estavam executando atividades de verificação de integridade, como coletar o estado do nó e o estado do YARN nó. HDFS Isso estava acontecendo porque os daemons no cluster não conseguiam comunicar os dados do status de saúde de um nó aos componentes internos da Amazon. EMR
- Daemons EMR no cluster aprimorados para rastrear corretamente os estados dos nós quando os endereços IP são reutilizados para melhorar a confiabilidade durante as operações de escalabilidade.
- [SPARK-29683](#). Corrigido um problema em que ocorriam falhas de trabalho durante a redução da escala verticalmente do cluster, pois o Spark presumia que todos os nós disponíveis estavam na lista de negação.
- [YARN-9011](#). Corrigido um problema em que falhas de trabalho ocorriam devido a uma condição de corrida no YARN descomissionamento quando o cluster tentava aumentar ou diminuir a escala.
- Foi corrigido o problema com falhas de etapas ou tarefas durante a escalabilidade do cluster, garantindo que os estados dos nós fossem sempre consistentes entre os EMR daemons no cluster da Amazon e/. YARN HDFS
- Corrigido um problema em que operações de cluster, como redução de escala e envio de etapas, falhavam em EMR clusters da Amazon habilitados com a autenticação Kerberos. Isso ocorreu porque o daemon no EMR cluster da Amazon não renovou o tíquete Kerberos, que é necessário para se comunicar de forma segura com/em execução no nó primário. HDFS YARN
- As EMR versões mais recentes da Amazon corrigem o problema com um limite menor de “Máximo de arquivos abertos” em relação aos mais antigos AL2 na AmazonEMR. As EMR versões 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 e posteriores agora incluem uma correção permanente com uma configuração mais alta de “Máximo de arquivos abertos”.
- Configurando um cluster para corrigir problemas de desempenho do Apache YARN Timeline Server versões 1 e 1.5

As versões 1 e 1.5 do Apache YARN Timeline Server podem causar problemas de desempenho com EMR clusters grandes e muito ativos, especialmente `comyarn.resourcemanager.system-metrics-publisher.enabled=true`, que é a configuração padrão na Amazon. EMR Um YARN Timeline Server v2 de código aberto resolve o problema de desempenho relacionado à escalabilidade do YARN Timeline Server.

Outras soluções alternativas para esse problema incluem:

- Configurando `yarn.resourcemanager.system-metrics-publisher.enabled=false` em `yarn-site.xml`.

- Habilitar a correção para esse problema na criação de um cluster, conforme descrito abaixo.

As seguintes EMR versões da Amazon contêm uma correção para esse problema de desempenho YARN do Timeline Server.

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

Para habilitar a correção em qualquer uma das EMR versões da Amazon especificadas acima, defina essas propriedades `true` em um JSON arquivo de configurações que é passado usando o [parâmetro de aws emr create-cluster comando](#): `--configurations file:///./configurations.json`. Ou habilite a correção usando a [interface do usuário do console de reconfiguração](#).

Exemplo de conteúdo do arquivo `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": []
  }
]
```

- O runtime do Spark agora é mais rápido na busca de localizações de partições do Hive Metastore para consultas de inserção do Spark.
- Atualizadas versões do componente. Para obter uma lista das versões dos componentes, consulte [Sobre os EMR lançamentos da Amazon](#) neste guia.
- Instalou o AWS Java SDK Bundle em cada novo cluster. Esse é um único jar contendo todos os serviços SDKs e suas dependências, em vez de jars de componentes individuais. Para obter mais informações, consulte [Java SDK Bundled Dependency](#).
- Corrigiu problemas de escalabilidade gerenciada em EMR versões anteriores da Amazon e fez melhorias para que as taxas de falha de aplicativos fossem significativamente reduzidas.
- HTTPS agora está habilitado por padrão para repositórios Amazon Linux. Se você estiver usando uma VPC política do Amazon S3 para restringir o acesso a buckets específicos, deverá adicionar o novo bucket Amazon Linux ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` à

sua política (\$regionsubstitua pela região onde está o endpoint). Para obter mais informações, consulte esse tópico nos fóruns de AWS discussão. [Anúncio: o Amazon Linux 2 agora oferece suporte à capacidade de uso HTTPS ao se conectar a repositórios de pacotes.](#)

Novos atributos

- A Amazon EMR oferece suporte aos pontos de acesso do Amazon S3, um recurso do Amazon S3 que permite gerenciar facilmente o acesso a data lakes compartilhados. Usando o alias do Amazon S3 Access Point, você pode simplificar o acesso aos dados em grande escala na Amazon. EMR Você pode usar os pontos de acesso do Amazon S3 com todas as versões da Amazon sem EMR custo adicional em todas as AWS regiões onde a Amazon EMR está disponível. Para saber mais sobre Pontos de Acesso Amazon S3 e aliases de ponto de acesso, consulte [Usar um alias em estilo de bucket para seu ponto de acesso](#) no Guia do usuário do Amazon S3.
- O Amazon EMR -5.33 suporta novos tipos de EC2 instância da Amazon: c5a, c5ad, c6gn, c6gd, m6gd, d3, d3en, m5zn, r5b, r6gd. Consulte [Tipos de instância compatíveis.](#)

Problemas conhecidos

- Limite inferior de “Máximo de arquivos abertos” em versões mais antigas AL2 [corrigido em versões mais recentes]. EMRLançamentos da Amazon: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 e emr-6.2.0 são baseados em versões mais antigas do Linux ofAmazon 2 (AL2, que têm uma configuração de limite inferior para “Máximo de arquivos abertos” quando os clusters da Amazon são criados com o padrão. EMR AMI As EMR versões 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 e posteriores incluem uma correção permanente com uma configuração maior de “Máximo de arquivos abertos”. Versões com o limite inferior de arquivos abertos causam o erro “Muitos arquivos abertos” ao ser enviado um trabalho do Spark. Nas versões afetadas, o padrão da Amazon AMI tem uma configuração EMR padrão de ulimit de 4096 para “Máximo de arquivos abertos”, que é inferior ao limite de 65536 arquivos no Linux 2. latestAmazon AMI A configuração inferior de ulimit para “Máximo de arquivos abertos” causa falhas em trabalhos do Spark quando o driver e o executor do Spark tentam abrir mais de 4096 arquivos. Para corrigir o problema, a Amazon EMR tem um script de ação de bootstrap (BA) que ajusta a configuração ulimit na criação do cluster.

Se você estiver usando uma EMR versão mais antiga da Amazon que não tem a solução permanente para esse problema, a solução alternativa a seguir permite que você defina explicitamente o ulimit do controlador de instância para um máximo de 65536 arquivos.

Defina explicitamente um ulimit na linha de comando

1. Edite `/etc/systemd/system/instance-controller.service` para adicionar os seguintes parâmetros à seção Serviço.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Reiniciar InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Defina um ulimit usando a ação de bootstrap (BA)

Você também pode usar um script de ação de bootstrap (BA) para configurar o ulimit instance-controller para 65536 arquivos na criação do 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
```

- Para clusters de sub-rede privados Amazon EMR 6.3.0 e 6.2.0, você não pode acessar a interface web do Ganglia. Você receberá um erro de “acesso negado (403)”. Outros sites UIs, como Spark, Hue, Zeppelin JupyterHub, Livy e Tez, estão funcionando normalmente. O acesso à interface

do usuário da Web do Ganglia em clusters de sub-redes públicas também está funcionando normalmente. Para resolver esse problema, reinicie o serviço `httpd` no nó primário com `sudo systemctl restart httpd`. Esse problema foi corrigido no Amazon EMR 6.4.0.

• **⚠ Important**

EMRclusters que executam Amazon Linux ou Amazon Linux 2 Amazon Machine Images (AMIs) usam o comportamento padrão do Amazon Linux e não baixam e instalam automaticamente atualizações importantes e críticas do kernel que exigem uma reinicialização. Esse é o mesmo comportamento de outras EC2 instâncias da Amazon que executam o Amazon Linux padrãoAMI. Se novas atualizações de software Amazon Linux que exigem uma reinicialização (como kernel e CUDA atualizações) ficarem disponíveis após a disponibilização de uma EMR versão da Amazon, as instâncias de EMR cluster que executam o padrão AMI não baixam e instalam automaticamente essas atualizações. NVIDIA Para obter atualizações do kernel, você pode [personalizar sua Amazon EMR AMI](#) para [usar o Amazon Linux AMI mais recente](#).

- O suporte do console para criar uma configuração de segurança que especifica a opção de integração do AWS Ranger atualmente não é suportado na GovCloud região. A configuração de segurança pode ser feita usando CLI o. Consulte [Criar a configuração EMR de segurança](#) no Guia EMR de gerenciamento da Amazon.
- Políticas gerenciadas com escopo definido: para se alinhar às AWS melhores práticas, a Amazon EMR introduziu políticas EMR gerenciadas padrão com escopo v2 como substitutas das políticas que serão descontinuadas. Consulte as [políticas EMR gerenciadas da Amazon](#).
- Quando você usa o Spark com a formatação de localização de partições do Hive para ler dados no Amazon S3 e executa o Spark nas EMR versões 5.30.0 a 5.36.0 e 6.2.0 a 6.9.0 da Amazon, você pode encontrar um problema que impede seu cluster de ler os dados corretamente. Isso poderá acontecer se suas partições tiverem todas as características a seguir:
 - Duas ou mais partições são verificadas na mesma tabela.
 - Pelo menos um caminho de diretório de partição é um prefixo de pelo menos outro caminho de diretório de partição, por exemplo, `s3://bucket/table/p=a` é um prefixo de `s3://bucket/table/p=a b`.
 - O primeiro caractere que segue o prefixo no outro diretório de partição tem um valor UTF -8 menor que o / caractere (U+002F). Por exemplo, o caractere de espaço (U+0020) que ocorre entre a e b em `s3://bucket/table/p=a b` se enquadra nessa categoria. Observe que

existem 14 outros caracteres que não são de controle: !"#%&'()*+,-. Para obter mais informações, consulte a [tabela de codificação UTF -8 e caracteres Unicode](#).

Como solução alternativa para esse problema, defina a configuração `spark.sql.sources.fastS3PartitionDiscovery.enabled` como `false` na classificação `spark-defaults`.

Versões de componente 5.33.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.16.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.14.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.5.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
emr-notebook-env	1.2.0	Ambiente Conda para bloco de anotações do emr, que inclui o jupyter enterprise gateway
emr-s3-dist-cp	2.18.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	1.6.0	EMRConector S3Select
emrfs	2.46.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.12.1	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
flink-jobmanager-config	1.12.1	Gerenciando recursos em EMR nós para o Apache JobManager Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.

Componente	Version (Versão)	Descrição
hadoop-client	2.10.1-amzn-1.1	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.10.1-amzn-1.1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.10.1-amzn-1.1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.10.1-amzn-1.1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	2.10.1-amzn-1.1	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httpfs-server	2.10.1-amzn-1.1	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.10.1-amzn-1.1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.10.1-amzn-1.1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.10.1-amzn-1.1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.10.1-amzn-1.1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.

Componente	Version (Versão)	Descrição
hadoop-yarn-timeline-server	2.10.1-amzn-1.1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.13	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.13	Serviço para atender uma ou mais HBase regiões.
hbase-client	1.4.13	HBase cliente de linha de comando.
hbase-rest-server	1.4.13	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.4.13	Serviço que fornece um endpoint Thrift para HBase.
hcatalog-client	2.3.7-amzn-4	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.7-amzn-4	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.7-amzn-4	HTTP endpoint fornecendo uma REST interface para HCatalog.

Componente	Version (Versão)	Descrição
hive-client	2.3.7-amzn-4	O cliente da linha de comando do Hive.
hive-hbase	2.3.7-amzn-4	Cliente do Hive-hbase.
hive-metastore-server	2.3.7-amzn-4	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.7-amzn-4	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.7.0-amzn-1	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.
hudi-spark	0.7.0-amzn-1	Biblioteca de pacotes para executar o Spark com o Hudi.
hudi-presto	0.7.0-amzn-1	Biblioteca de pacotes para executar o Presto com o Hudi.
hue-server	4.9.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	1.2.2	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.7.0-incubating	RESTinterface para interagir com o Apache Spark

Componente	Version (Versão)	Descrição
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.7.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mariadb-server	5.5.68+	Meu servidor SQL de banco de dados.
nvidia-cuda	10.1.243	Drivers NVIDIA e toolkit CUDA
oozie-client	5.2.0	O cliente da linha de comando do Oozie.
oozie-server	5.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	4.5.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.3- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.3- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.245.1-amzn-0	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.

Componente	Version (Versão)	Descrição
presto-worker	0.245.1-amzn-0	O serviço que executa partes de uma consulta.
presto-client	0.245.1-amzn-0	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	4.0.2	O projeto R para computação estatística
ranger-kms-server	1.2.0	Sistema de gerenciamento de chaves Apache Ranger
spark-client	2.4.7-amzn-1.1	Os clientes da linha de comando do Spark.
spark-history-server	2.4.7-amzn-1.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.4.7-amzn-1.1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.4.7-amzn-1.1	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.

Componente	Version (Versão)	Descrição
tensorflow	2.4.1	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.2	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.9.0	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.14	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.14	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.33.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Ocorrem ações de reconfiguração quando você especifica uma configuração para grupos de instâncias em um cluster em execução. A Amazon EMR só inicia ações de reconfiguração para as classificações que você modifica. Para ter mais informações, consulte [Reconfigurar um grupo de instâncias de um cluster em execução](#).

Classificações do emr-5.33.1

Classificações	Descrição	Ações de reconfiguração
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.	Restarts the ResourceManager service.
container-executor	Altere os valores no arquivo container-executor.cfg YARN do Hadoop.	Not available.
container-log4j	Altere os valores no arquivo container-log4j.properties YARN do Hadoop.	Not available.
core-site	Alterar os valores no arquivo core-site.xml do 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	Alterar as configurações relacionadas ao docker.	Not available.
emrfs-site	Altere EMRFS as configurações.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the

Classificações	Descrição	Ações de reconfiguração
		Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts HBaseRegionserver, HBaseMaster, HBaseThrift, HBaseRest, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
flink-conf	Alterar as configurações do flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Alterar as configurações de log4j.properties no Flink.	Restarts Flink history server.
flink-log4j-session	Alterar as configurações de log4j-yarn-session.properties no Flink para sessão do Kubernetes/Yarn.	Not available.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.	Restarts Flink history server.

Classificações	Descrição	Ações de reconfiguração
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do 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	Alterar os valores no arquivo log4j.properties do 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	Alterar a configuração do servidor SSL no Hadoop	Not available.
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop	Not available.

Classificações	Descrição	Ações de reconfiguração
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Mude os valores no HBase ambiente.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.	Not available.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.	Should not be reconfigured.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Classificações	Descrição	Ações de reconfiguração
hcatalog-env	Mude os valores no HCatalog ambiente.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.	Not available.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.	Not available.
hive-env	Alterar os valores no ambiente do Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log 4j2.properties do Hive.	Restarts HiveServer2 and HiveMetastore.

Classificações	Descrição	Ações de reconfiguração
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log 4j2.properties do Hive.	Not available.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.	Not available.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.	Not available.
hue-ini	Alterar os valores no arquivo ini do Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Mude os valores no HTTPFS ambiente.	Restarts Hadoop Httpfs service.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.	Not available.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.	Not available.

Classificações	Descrição	Ações de reconfiguração
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Altere os valores no ambiente do Hudi.	Not available.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.	Not available.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.	Not available.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.	Not available.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.	Not available.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.	Restarts Livy Server.
livy-env	Alterar os valores no ambiente do Livy.	Restarts Livy Server.
livy-log4j	Alterar as configurações de log4j.properties no Livy.	Restarts Livy Server.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.

Classificações	Descrição	Ações de reconfiguração
oozie-env	Alterar os valores no ambiente do Oozie.	Restarts Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.	Restarts Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.	Not available.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.	Not available.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.	Not available.
pig-env	Alterar os valores no ambiente do Pig.	Not available.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.	Restarts Oozie.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.	Not available.
presto-log	Alterar os valores no arquivo log.properties do Presto.	Restarts Presto-Server.

Classificações	Descrição	Ações de reconfiguração
presto-config	Alterar os valores no arquivo <code>config.properties</code> do Presto.	Restarts Presto-Server.
presto-password-authenticator	Altere os valores no arquivo <code>password-authenticator.properties</code> do Presto.	Not available.
presto-env	Altere os valores no arquivo <code>presto-env.sh</code> do Presto.	Restarts Presto-Server.
presto-node	Altere os valores no arquivo <code>node.properties</code> do Presto.	Not available.
presto-connector-blackhole	Alterar os valores no arquivo <code>blackhole.properties</code> do Presto.	Not available.
presto-connector-cassandra	Alterar os valores no arquivo <code>cassandra.properties</code> do Presto.	Not available.
presto-connector-hive	Alterar os valores no arquivo <code>hive.properties</code> do Presto.	Restarts Presto-Server.
presto-connector-jmx	Alterar os valores no arquivo <code>jmx.properties</code> do Presto.	Not available.
presto-connector-kafka	Alterar os valores no arquivo <code>kafka.properties</code> do Presto.	Not available.
presto-connector-localfile	Alterar os valores no arquivo <code>localfile.properties</code> do Presto.	Not available.
presto-connector-memory	Alterar os valores no arquivo <code>memory.properties</code> do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-mongodb	Alterar os valores no arquivo <code>mongodb.properties</code> do Presto.	Not available.
presto-connector-mysql	Alterar os valores no arquivo <code>mysql.properties</code> do Presto.	Not available.
presto-connector-postgresql	Alterar os valores no arquivo <code>postgresql.properties</code> do Presto.	Not available.
presto-connector-raptor	Alterar os valores no arquivo <code>raptor.properties</code> do Presto.	Not available.
presto-connector-redis	Alterar os valores no arquivo <code>redis.properties</code> do Presto.	Not available.
presto-connector-redshift	Alterar os valores no arquivo de propriedades <code>redshift</code> do Presto.	Not available.
presto-connector-tpch	Alterar os valores no arquivo <code>tpch.properties</code> do Presto.	Not available.
presto-connector-tpcds	Alterar os valores no arquivo <code>tpcds.properties</code> do Presto.	Not available.
ranger-kms-dbks-site	Altere os valores no arquivo <code>dbks-site.xml</code> do Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-site	Altere os valores no <code>ranger-kms-site</code> arquivo.xml do RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.	Restarts Ranger KMS Server.

Classificações	Descrição	Ações de reconfiguração
ranger-kms-log4j	Altere os valores no arquivo kms-log4j.properties do Ranger. KMS	Not available.
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.	Not available.
recordserver-env	Mude os valores no EMR RecordServer ambiente.	Restarts EMR record server.
recordserver-conf	Altere os valores no arquivo EMR RecordServer erver.properties do.	Restarts EMR record server.
recordserver-log4j	Altere os valores no EMR RecordServer arquivo log4j.properties.	Restarts EMR record server.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Alterar os valores no ambiente do Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.	Not available.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.	Restarts Spark history server and Spark thrift server.

Classificações	Descrição	Ações de reconfiguração
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Alterar os valores no ambiente do Sqoop.	Not available.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.	Not available.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.	Not available.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.	Restarts Oozie.
yarn-env	Mude os valores no YARN ambiente.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.	Restarts Zeppelin.
zeppelin-site	Alterar as definições da configuração em zeppelin-site.xml.	Restarts Zeppelin.

Classificações	Descrição	Ações de reconfiguração
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.	Restarts Zookeeper server.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.	Restarts Zookeeper server.

Amazon EMR versão 5.33.0

Versões 5.33.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.33.0	emr-5.32.1	emr-5.32.0	emr-5.31.1
AWS SDK para Java	1.11.970	1.11.890	1.11.890	1.11.852
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.11.12	2.11.12	2.11.12	2.11.12

	emr-5.33.0	emr-5.32.1	emr-5.32.0	emr-5.31.1
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
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

	emr-5.33.0	emr-5.32.1	emr-5.32.0	emr-5.31.1
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
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

Notas da versão 5.33.0

Versões de componente 5.33.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK
emr-ddb	4.16.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.14.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.5.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-notebook-env	1.2.0	Ambiente Conda para bloco de anotações do emr, que inclui o jupyter enterprise gateway
emr-s3-dist-cp	2.18.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	1.6.0	EMRConector S3Select
emrfs	2.46.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.12.1	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
flink-jobmanager-config	1.12.1	Gerenciando recursos em EMR nós para o Apache JobManager Flink.

Componente	Version (Versão)	Descrição
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.10.1-amzn-1	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.10.1-amzn-1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.10.1-amzn-1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.10.1-amzn-1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	2.10.1-amzn-1	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httpfs-server	2.10.1-amzn-1	HTTPendpoint para HDFS operações.

Componente	Version (Versão)	Descrição
hadoop-kms-server	2.10.1-amzn-1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.10.1-amzn-1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.10.1-amzn-1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.10.1-amzn-1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.10.1-amzn-1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.13	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.13	Serviço para atender uma ou mais HBase regiões.
hbase-client	1.4.13	HBasecliente de linha de comando.
hbase-rest-server	1.4.13	Serviço que fornece um RESTful HTTP endpoint paraHBase.

Componente	Version (Versão)	Descrição
hbase-thrift-server	1.4.13	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	2.3.7-amzn-4	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.7-amzn-4	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.7-amzn-4	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	2.3.7-amzn-4	O cliente da linha de comando do Hive.
hive-hbase	2.3.7-amzn-4	Cliente do Hive-hbase.
hive-metastore-server	2.3.7-amzn-4	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.7-amzn-4	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.7.0-amzn-1	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.

Componente	Version (Versão)	Descrição
hudi-spark	0.7.0-amzn-1	Biblioteca de pacotes para executar o Spark com o Hudi.
hudi-presto	0.7.0-amzn-1	Biblioteca de pacotes para executar o Presto com o Hudi.
hue-server	4.9.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	1.2.2	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.7.0-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.7.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mariadb-server	5.5.68	Meu servidor SQL de banco de dados.
nvidia-cuda	10.1.243	Drivers NVIDIA e toolkit CUDA
oozie-client	5.2.0	O cliente da linha de comando do Oozie.

Componente	Version (Versão)	Descrição
oozie-server	5.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	4.5.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.3- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.3- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.245.1-amzn-0	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.245.1-amzn-0	O serviço que executa partes de uma consulta.
presto-client	0.245.1-amzn-0	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	4.0.2	O projeto R para computação estatística

Componente	Version (Versão)	Descrição
ranger-kms-server	1.2.0	Sistema de gerenciamento de chaves Apache Ranger
spark-client	2.4.7-amzn-1	Os clientes da linha de comando do Spark.
spark-history-server	2.4.7-amzn-1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.4.7-amzn-1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.4.7-amzn-1	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	2.4.1	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.2	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.9.0	O notebook baseado na web que permite a análise de dados interativa.

Componente	Version (Versão)	Descrição
zookeeper-server	3.4.14	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.14	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.33.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Ocorrem ações de reconfiguração quando você especifica uma configuração para grupos de instâncias em um cluster em execução. A Amazon EMR só inicia ações de reconfiguração para as classificações que você modifica. Para ter mais informações, consulte [Reconfigurar um grupo de instâncias de um cluster em execução](#).

Classificações do emr-5.33.0

Classificações	Descrição	Ações de reconfiguração
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.	Restarts the Resource Manager service.
container-executor	Altere os valores no arquivo <code>container-executor.cfg</code> YARN do Hadoop.	Not available.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.	Not available.

Classificações	Descrição	Ações de reconfiguração
core-site	Alterar os valores no arquivo core-site.xml do 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	Alterar as configurações relacionadas ao docker.	Not available.
emrfs-site	Altere EMRFS as configurações.	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.

Classificações	Descrição	Ações de reconfiguração
flink-conf	Alterar as configurações do flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Alterar as configurações de log4j.properties no Flink.	Restarts Flink history server.
flink-log4j-session	Alterar as configurações de log4j-yarn-session.properties no Flink para sessão do Kubernetes/Yarn.	Not available.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.	Restarts Flink history server.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do 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.

Classificações	Descrição	Ações de reconfiguração
hadoop-log4j	Alterar os valores no arquivo log4j.properties do 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	Alterar a configuração do servidor SSL no Hadoop	Not available.
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop	Not available.
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Mude os valores no HBase ambiente.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.

Classificações	Descrição	Ações de reconfiguração
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.	Not available.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.	Should not be reconfigured.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Mude os valores no HCatalog ambiente.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat Server.

Classificações	Descrição	Ações de reconfiguração
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.	Not available.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.	Not available.
hive-env	Alterar os valores no ambiente do Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log 4j2.properties do Hive.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log 4j2.properties do Hive.	Not available.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.	Not available.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.	Not available.

Classificações	Descrição	Ações de reconfiguração
hue-ini	Alterar os valores no arquivo ini do Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Mude os valores no HTTPFS ambiente.	Restarts Hadoop Httpfs service.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.	Not available.
hadoop-kms-env	Altere os valores no ambiente do HadoopKMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.	Not available.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Altere os valores no ambiente do Hudi.	Not available.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.	Not available.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.	Not available.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.	Not available.

Classificações	Descrição	Ações de reconfiguração
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.	Not available.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.	Restarts Livy Server.
livy-env	Alterar os valores no ambiente do Livy.	Restarts Livy Server.
livy-log4j	Alterar as configurações de log4j.properties no Livy.	Restarts Livy Server.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Alterar os valores no ambiente do Oozie.	Restarts Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.	Restarts Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.	Not available.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.	Not available.

Classificações	Descrição	Ações de reconfiguração
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.	Not available.
pig-env	Alterar os valores no ambiente do Pig.	Not available.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.	Restarts Oozie.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.	Not available.
presto-log	Alterar os valores no arquivo log.properties do Presto.	Restarts Presto-Server.
presto-config	Alterar os valores no arquivo config.properties do Presto.	Restarts Presto-Server.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.	Not available.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.	Restarts Presto-Server.
presto-node	Altere os valores no arquivo node.properties do Presto.	Not available.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-cassandra	Alterar os valores no arquivo <code>cassandra.properties</code> do Presto.	Not available.
presto-connector-hive	Alterar os valores no arquivo <code>hive.properties</code> do Presto.	Restarts Presto-Server.
presto-connector-jmx	Alterar os valores no arquivo <code>jmx.properties</code> do Presto.	Not available.
presto-connector-kafka	Alterar os valores no arquivo <code>kafka.properties</code> do Presto.	Not available.
presto-connector-localfile	Alterar os valores no arquivo <code>localfile.properties</code> do Presto.	Not available.
presto-connector-memory	Alterar os valores no arquivo <code>memory.properties</code> do Presto.	Not available.
presto-connector-mongodb	Alterar os valores no arquivo <code>mongodb.properties</code> do Presto.	Not available.
presto-connector-mysql	Alterar os valores no arquivo <code>mysql.properties</code> do Presto.	Not available.
presto-connector-postgresql	Alterar os valores no arquivo <code>postgresql.properties</code> do Presto.	Not available.
presto-connector-raptor	Alterar os valores no arquivo <code>raptor.properties</code> do Presto.	Not available.
presto-connector-redis	Alterar os valores no arquivo <code>redis.properties</code> do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.	Not available.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.	Not available.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.	Not available.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Altere os valores no arquivo kms-log4j.properties do Ranger. KMS	Not available.
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.	Not available.
recordserver-env	Mude os valores no EMR RecordServer ambiente.	Restarts EMR record server.
recordserver-conf	Altere os valores no arquivo EMR RecordServer erver.properties do.	Restarts EMR record server.

Classificações	Descrição	Ações de reconfiguração
recordserver-log4j	Altere os valores no EMR RecordServer arquivo log4j.properties.	Restarts EMR record server.
spark	Configurações EMR seleciona das pela Amazon para o Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Alterar os valores no ambiente do Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.	Not available.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Alterar os valores no ambiente do Sqoop.	Not available.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.	Not available.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.	Not available.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.	Restarts Oozie.

Classificações	Descrição	Ações de reconfiguração
yarn-env	Mude os valores no YARN ambiente.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.	Restarts Zeppelin.
zeppelin-site	Alterar as definições da configuração em zeppelin-site.xml.	Restarts Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.	Restarts Zookeeper server.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.	Restarts Zookeeper server.

Amazon EMR versão 5.32.1

Versões 5.32.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.32.1	emr-5.32.0	emr-5.31.1	emr-5.31.0
AWS SDK para Java	1.11.890	1.11.890	1.11.852	1.11.852
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.11.12	2.11.12	2.11.12	2.11.12
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

	emr-5.32.1	emr-5.32.0	emr-5.31.1	emr-5.31.0
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
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

Notas da versão 5.32.1

Esta é uma versão para corrigir problemas com o Amazon EMR Scaling quando ele falha na escalabilidade ou redução de escala de um cluster com sucesso ou causa falhas no aplicativo.

Alterações, melhorias e problemas resolvidos

- Corrigido um problema em que as solicitações de escalabilidade falhavam em um cluster grande e altamente utilizado quando os daemons no cluster da EMR Amazon estavam executando atividades de verificação de integridade, como coletar o estado do nó e o estado do YARN nó. HDFS Isso estava acontecendo porque os daemons no cluster não conseguiam comunicar os dados do status de saúde de um nó aos componentes internos da Amazon. EMR
- Daemons EMR no cluster aprimorados para rastrear corretamente os estados dos nós quando os endereços IP são reutilizados para melhorar a confiabilidade durante as operações de escalabilidade.
- [SPARK-29683](#). Corrigido um problema em que ocorriam falhas de trabalho durante a redução da escala verticalmente do cluster, pois o Spark presumia que todos os nós disponíveis estavam na lista de negação.
- [YARN-9011](#). Corrigido um problema em que falhas de trabalho ocorriam devido a uma condição de corrida no YARN descomissionamento quando o cluster tentava aumentar ou diminuir a escala.
- Foi corrigido o problema com falhas de etapas ou tarefas durante a escalabilidade do cluster, garantindo que os estados dos nós fossem sempre consistentes entre os EMR daemons no cluster da Amazon e/. YARN HDFS
- Corrigido um problema em que operações de cluster, como redução de escala e envio de etapas, falhavam em EMR clusters da Amazon habilitados com a autenticação Kerberos. Isso ocorreu porque o daemon no EMR cluster da Amazon não renovou o tíquete Kerberos, que é necessário para se comunicar de forma segura com/em execução no nó primário. HDFS YARN
- As EMR versões mais recentes da Amazon corrigem o problema com um limite menor de “Máximo de arquivos abertos” em relação aos mais antigos AL2 na AmazonEMR. As EMR versões 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 e posteriores agora incluem uma correção permanente com uma configuração mais alta de “Máximo de arquivos abertos”.
- HTTPS agora está habilitado por padrão para repositórios Amazon Linux. Se você estiver usando uma VPC política do Amazon S3 para restringir o acesso a buckets específicos, deverá adicionar o novo bucket Amazon Linux ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` à sua política (`$region` substitua pela região onde está o endpoint). Para obter mais informações,

consulte esse tópico nos fóruns de AWS discussão. [Anúncio: o Amazon Linux 2 agora oferece suporte à capacidade de uso HTTPS ao se conectar a repositórios de pacotes.](#)

Problemas conhecidos

- Quando você usa o Spark com a formatação de localização de partições do Hive para ler dados no Amazon S3 e executa o Spark nas EMR versões 5.30.0 a 5.36.0 e 6.2.0 a 6.9.0 da Amazon, você pode encontrar um problema que impede seu cluster de ler os dados corretamente. Isso poderá acontecer se suas partições tiverem todas as características a seguir:
 - Duas ou mais partições são verificadas na mesma tabela.
 - Pelo menos um caminho de diretório de partição é um prefixo de pelo menos outro caminho de diretório de partição, por exemplo, `s3://bucket/table/p=a` é um prefixo de `s3://bucket/table/p=a b`.
 - O primeiro caractere que segue o prefixo no outro diretório de partição tem um valor UTF -8 menor que o / caractere (U+002F). Por exemplo, o caractere de espaço (U+0020) que ocorre entre a e b em `s3://bucket/table/p=a b` se enquadra nessa categoria. Observe que existem 14 outros caracteres que não são de controle: `!"#$%&'()*+,-.` Para obter mais informações, consulte a [tabela de codificação UTF -8 e caracteres Unicode](#).

Como solução alternativa para esse problema, defina a configuração `spark.sql.sources.fastS3PartitionDiscovery.enabled` como `false` na classificação `spark-defaults`.

Versões de componente 5.32.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
aws-sagemaker-spark-sdk	1.4.1	Amazon SageMaker Spark SDK
emr-ddb	4.16.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.13.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.5.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-notebook-env	1.1.0	Ambiente Conda para bloco de anotações do emr, que inclui o jupyter enterprise gateway
emr-s3-dist-cp	2.17.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	1.6.0	EMRConector S3Select
emrfs	2.45.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.11.2	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
flink-jobmanager-config	1.11.2	Gerenciando recursos em EMR nós para o Apache JobManager Flink.

Componente	Version (Versão)	Descrição
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.10.1-amzn-0.1	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.10.1-amzn-0.1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.10.1-amzn-0.1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.10.1-amzn-0.1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	2.10.1-amzn-0.1	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httpfs-server	2.10.1-amzn-0.1	HTTPendpoint para HDFS operações.

Componente	Version (Versão)	Descrição
hadoop-kms-server	2.10.1-amzn-0.1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.10.1-amzn-0.1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.10.1-amzn-0.1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.10.1-amzn-0.1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.10.1-amzn-0.1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.13	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.13	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.4.13	HBasecliente de linha de comando.
hbase-rest-server	1.4.13	Serviço que fornece um RESTful HTTP endpoint paraHBase.

Componente	Version (Versão)	Descrição
hbase-thrift-server	1.4.13	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	2.3.7-amzn-3	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.7-amzn-3	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.7-amzn-3	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	2.3.7-amzn-3	O cliente da linha de comando do Hive.
hive-hbase	2.3.7-amzn-3	Cliente do Hive-hbase.
hive-metastore-server	2.3.7-amzn-3	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.7-amzn-3	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.6.0-amzn-0	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.

Componente	Version (Versão)	Descrição
hudi-spark	0.6.0-amzn-0	Biblioteca de pacotes para executar o Spark com o Hudi.
hudi-presto	0.6.0-amzn-0	Biblioteca de pacotes para executar o Presto com o Hudi.
hue-server	4.8.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	1.1.0	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.7.0-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.7.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mariadb-server	5.5.68+	Meu servidor SQL de banco de dados.
nvidia-cuda	10.1.243	Drivers NVIDIA e toolkit CUDA
oozie-client	5.2.0	O cliente da linha de comando do Oozie.

Componente	Version (Versão)	Descrição
oozie-server	5.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	4.4.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.3- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.3- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.240.1-amzn-0	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.240.1-amzn-0	O serviço que executa partes de uma consulta.
presto-client	0.240.1-amzn-0	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.3	O projeto R para computação estatística

Componente	Version (Versão)	Descrição
ranger-kms-server	1.2.0	Sistema de gerenciamento de chaves Apache Ranger
spark-client	2.4.7-amzn-0.1	Os clientes da linha de comando do Spark.
spark-history-server	2.4.7-amzn-0.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.4.7-amzn-0.1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.4.7-amzn-0.1	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	2.3.1	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.2	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.
zeppelin-server	0.8.2	O notebook baseado na web que permite a análise de dados interativa.

Componente	Version (Versão)	Descrição
zookeeper-server	3.4.14	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.14	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.32.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Ocorrem ações de reconfiguração quando você especifica uma configuração para grupos de instâncias em um cluster em execução. A Amazon EMR só inicia ações de reconfiguração para as classificações que você modifica. Para ter mais informações, consulte [Reconfigurar um grupo de instâncias de um cluster em execução](#).

Classificações do emr-5.32.1

Classificações	Descrição	Ações de reconfiguração
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.	Restarts the Resource Manager service.
container-executor	Altere os valores no arquivo <code>container-executor.cfg</code> YARN do Hadoop.	Not available.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.	Not available.

Classificações	Descrição	Ações de reconfiguração
core-site	Alterar os valores no arquivo core-site.xml do 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	Alterar as configurações relacionadas ao docker.	Not available.
emrfs-site	Altere EMRFS as configurações.	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.

Classificações	Descrição	Ações de reconfiguração
flink-conf	Alterar as configurações do flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Alterar as configurações de log4j.properties no Flink.	Restarts Flink history server.
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.	Not available.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.	Restarts Flink history server.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do 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.

Classificações	Descrição	Ações de reconfiguração
hadoop-log4j	Alterar os valores no arquivo log4j.properties do 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	Alterar a configuração do servidor SSL no Hadoop	Not available.
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop	Not available.
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Mude os valores no HBase ambiente.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.

Classificações	Descrição	Ações de reconfiguração
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.	Not available.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.	Should not be reconfigured.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.
hcatalog-env	Mude os valores no HCatalog ambiente.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat Server.

Classificações	Descrição	Ações de reconfiguração
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.	Not available.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.	Not available.
hive-env	Alterar os valores no ambiente do Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log 4j2.properties do Hive.	Restarts HiveServer2 and HiveMetastore.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log 4j2.properties do Hive.	Not available.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.	Not available.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.	Not available.

Classificações	Descrição	Ações de reconfiguração
hue-ini	Alterar os valores no arquivo ini do Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Mude os valores no HTTPFS ambiente.	Restarts Hadoop Httpfs service.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.	Not available.
hadoop-kms-env	Altere os valores no ambiente do HadoopKMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.	Not available.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Altere os valores no ambiente do Hudi.	Not available.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.	Not available.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.	Not available.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.	Not available.

Classificações	Descrição	Ações de reconfiguração
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.	Not available.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.	Restarts Livy Server.
livy-env	Alterar os valores no ambiente do Livy.	Restarts Livy Server.
livy-log4j	Alterar as configurações de log4j.properties no Livy.	Restarts Livy Server.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
oozie-env	Alterar os valores no ambiente do Oozie.	Restarts Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.	Restarts Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.	Not available.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.	Not available.

Classificações	Descrição	Ações de reconfiguração
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.	Not available.
pig-env	Alterar os valores no ambiente do Pig.	Not available.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.	Restarts Oozie.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.	Not available.
presto-log	Alterar os valores no arquivo log.properties do Presto.	Restarts Presto-Server.
presto-config	Alterar os valores no arquivo config.properties do Presto.	Restarts Presto-Server.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.	Not available.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.	Restarts Presto-Server.
presto-node	Altere os valores no arquivo node.properties do Presto.	Not available.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-cassandra	Alterar os valores no arquivo <code>cassandra.properties</code> do Presto.	Not available.
presto-connector-hive	Alterar os valores no arquivo <code>hive.properties</code> do Presto.	Restarts Presto-Server.
presto-connector-jmx	Alterar os valores no arquivo <code>jmx.properties</code> do Presto.	Not available.
presto-connector-kafka	Alterar os valores no arquivo <code>kafka.properties</code> do Presto.	Not available.
presto-connector-localfile	Alterar os valores no arquivo <code>localfile.properties</code> do Presto.	Not available.
presto-connector-memory	Alterar os valores no arquivo <code>memory.properties</code> do Presto.	Not available.
presto-connector-mongodb	Alterar os valores no arquivo <code>mongodb.properties</code> do Presto.	Not available.
presto-connector-mysql	Alterar os valores no arquivo <code>mysql.properties</code> do Presto.	Not available.
presto-connector-postgresql	Alterar os valores no arquivo <code>postgresql.properties</code> do Presto.	Not available.
presto-connector-raptor	Alterar os valores no arquivo <code>raptor.properties</code> do Presto.	Not available.
presto-connector-redis	Alterar os valores no arquivo <code>redis.properties</code> do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.	Not available.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.	Not available.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.	Not available.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.	Restarts Ranger KMS Server.
ranger-kms-log4j	Altere os valores no arquivo kms-log4j.properties do Ranger. KMS	Not available.
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.	Not available.
recordserver-env	Mude os valores no EMR RecordServer ambiente.	Restarts EMR record server.
recordserver-conf	Altere os valores no arquivo EMR RecordServer erver.properties do.	Restarts EMR record server.

Classificações	Descrição	Ações de reconfiguração
recordserver-log4j	Altere os valores no EMR RecordServer arquivo log4j.properties.	Restarts EMR record server.
spark	Configurações EMR seleciona das pela Amazon para o Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Alterar os valores no ambiente do Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.	Not available.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.	Restarts Spark history server and Spark thrift server.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Alterar os valores no ambiente do Sqoop.	Not available.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.	Not available.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.	Not available.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.	Restarts Oozie.

Classificações	Descrição	Ações de reconfiguração
yarn-env	Mude os valores no YARN ambiente.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.	Restarts Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.	Restarts Zookeeper server.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.	Restarts Zookeeper server.

Amazon EMR versão 5.32.0

Versões 5.32.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterEnterpriseGateway](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.32.0	emr-5.31.1	emr-5.31.0	emr-5.30.2
AWS SDK para Java	1.11.890	1.11.852	1.11.852	1.11.759
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.11.12	2.11.12	2.11.12	2.11.12
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
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

	emr-5.32.0	emr-5.31.1	emr-5.31.0	emr-5.30.2
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
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

Notas da versão 5.32.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.32.0 da Amazon. As alterações são referentes à versão 5.31.0.

Data da versão inicial: 8 de janeiro de 2021

Atualizações

- Atualizado o conector do Amazon Glue para a versão 1.14.0
- Amazon SageMaker Spark atualizado SDK para a versão 1.4.1
- Atualizado para AWS SDK for Java a versão 1.11.890
- Atualização do EMR DynamoDB Connector versão 4.16.0
- Atualizado EMRFS para a versão 2.45.0
- Métricas do EMR Log Analytics atualizadas para a versão 1.18.0
- EMR MetricsAndEventsApiGateway Cliente atualizado para a versão 1.5.0
- Servidor de EMR registros atualizado para a versão 1.8.0
- EMRS3 Dist CP atualizado para a versão 2.17.0
- Agente EMR secreto atualizado para a versão 1.7.0
- Atualizado o Flink para a versão 1.11.2
- Atualizado o Hadoop para a versão 2.10.1-amzn-0
- Atualizado o Hive para a versão 2.3.7-amzn-3
- Atualizado o Hue para a versão 4.8.0
- Atualizado o Mxnet para a versão 1.7.0
- Atualizado o OpenCV para a versão 4.4.0
- Atualizado o Presto para a versão 0.240.1-amzn-0
- Atualizado o Spark para a versão 2.4.7-amzn-0
- Atualizado TensorFlow para a versão 2.3.1

Alterações, melhorias e problemas resolvidos

- Esta é uma versão para corrigir problemas com o Amazon EMR Scaling quando ele falha na escalabilidade ou redução de escala de um cluster com sucesso ou causa falhas no aplicativo.

- Corrigido um problema em que as solicitações de escalabilidade falhavam em um cluster grande e altamente utilizado quando os daemons no cluster da EMR Amazon estavam executando atividades de verificação de integridade, como coletar o estado do nó e o estado do YARN nó. HDFS Isso estava acontecendo porque os daemons no cluster não conseguiam comunicar os dados do status de saúde de um nó aos componentes internos da Amazon. EMR
- Daemons EMR no cluster aprimorados para rastrear corretamente os estados dos nós quando os endereços IP são reutilizados para melhorar a confiabilidade durante as operações de escalabilidade.
- [SPARK-29683](#). Corrigido um problema em que ocorriam falhas de trabalho durante a redução da escala verticalmente do cluster, pois o Spark presumia que todos os nós disponíveis estavam na lista de negação.
- [YARN-9011](#). Corrigido um problema em que falhas de trabalho ocorriam devido a uma condição de corrida no YARN descomissionamento quando o cluster tentava aumentar ou diminuir a escala.
- Foi corrigido o problema com falhas de etapas ou tarefas durante a escalabilidade do cluster, garantindo que os estados dos nós fossem sempre consistentes entre os EMR daemons no cluster da Amazon e/. YARN HDFS
- Corrigido um problema em que operações de cluster, como redução de escala e envio de etapas, falhavam em EMR clusters da Amazon habilitados com a autenticação Kerberos. Isso ocorreu porque o daemon no EMR cluster da Amazon não renovou o tíquete Kerberos, que é necessário para se comunicar de forma segura com/em execução no nó primário. HDFS YARN
- As EMR versões mais recentes da Amazon corrigem o problema com um limite menor de “Máximo de arquivos abertos” em relação aos mais antigos AL2 na AmazonEMR. As EMR versões 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 e posteriores agora incluem uma correção permanente com uma configuração mais alta de “Máximo de arquivos abertos”.
- Atualizadas versões do componente.
- Para obter uma lista das versões dos componentes, consulte [Sobre os EMR lançamentos da Amazon](#) neste guia.

Novos atributos

- A partir do Amazon EMR 5.32.0 e 6.5.0, o dimensionamento dinâmico do executor para o Apache Spark está habilitado por padrão. Para ativar ou desativar esse atributo, você pode usar o parâmetro de configuração `spark.yarn.heterogeneousExecutors.enabled`.
- Status de suporte do Instance Metadata Service (IMDS) V2: os componentes Amazon EMR 5.23.1, 5.27.1 e 5.32 ou posteriores são usados para todas as chamadas. IMDSv2 IMDS Para IMDS

chamadas no código do seu aplicativo, você pode usar ambos IMDSv1 e IMDSv2, ou configurar o IMDS para uso somente IMDSv2 para aumentar a segurança. Para outras EMR versões 5.x, a desativação IMDSv1 causa falha na inicialização do cluster.

- A partir do Amazon EMR 5.32.0, você pode iniciar um cluster que se integra nativamente com o Apache Ranger. O Apache Ranger é uma estrutura de código aberto para habilitar, monitorar e gerenciar uma segurança de dados abrangente em toda a plataforma Hadoop. Para obter mais informações, consulte [Apache Ranger](#). Com a integração nativa, você pode trazer seu próprio Apache Ranger para impor um controle refinado de acesso aos dados na Amazon. EMR Consulte [Integrar a Amazon EMR com o Apache Ranger](#) no Guia de EMR lançamento da Amazon.
- A EMR versão 5.32.0 da Amazon oferece suporte à Amazon EMR on EKS. Para obter mais detalhes sobre como começar a usar o EMR on EKS, consulte [What is Amazon EMR on EKS](#).
- A EMR versão 5.32.0 da Amazon é compatível com o Amazon EMR Studio (versão prévia). Para obter mais detalhes sobre como começar a usar o EMR Studio, consulte [Amazon EMR Studio \(versão prévia\)](#).
- Políticas gerenciadas com escopo definido: para se alinhar às AWS melhores práticas, a Amazon EMR introduziu políticas EMR gerenciadas padrão com escopo v2 como substitutas das políticas que serão descontinuadas. Consulte as [políticas EMR gerenciadas da Amazon](#).

Problemas conhecidos

- Para clusters de sub-rede privados Amazon EMR 6.3.0 e 6.2.0, você não pode acessar a interface web do Ganglia. Você receberá um erro de “acesso negado (403)”. Outros sites UIs, como Spark, Hue, Zeppelin JupyterHub, Livy e Tez, estão funcionando normalmente. O acesso à interface do usuário da Web do Ganglia em clusters de sub-redes públicas também está funcionando normalmente. Para resolver esse problema, reinicie o serviço httpd no nó primário com `sudo systemctl restart httpd`. Esse problema foi corrigido no Amazon EMR 6.4.0.
- Limite inferior de “Máximo de arquivos abertos” em versões mais antigas AL2 [corrigido em versões mais recentes]. EMR Lançamentos da Amazon: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 e emr-6.2.0 são baseados em versões mais antigas do Linux of Amazon 2 (AL2), que têm uma configuração de limite inferior para “Máximo de arquivos abertos” quando os clusters da Amazon são criados com o padrão. EMR AMI As EMR versões 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 e posteriores incluem uma correção permanente com uma configuração maior de “Máximo de arquivos abertos”. Versões com o limite inferior de arquivos abertos causam o erro “Muitos arquivos abertos” ao ser enviado um trabalho do Spark. Nas versões afetadas, o padrão da Amazon AMI tem uma configuração EMR padrão de `ulimit`

de 4096 para “Máximo de arquivos abertos”, que é inferior ao limite de 65536 arquivos no Linux 2. latestAmazon AMI A configuração inferior de ulimit para “Máximo de arquivos abertos” causa falhas em trabalhos do Spark quando o driver e o executor do Spark tentam abrir mais de 4096 arquivos. Para corrigir o problema, a Amazon EMR tem um script de ação de bootstrap (BA) que ajusta a configuração ulimit na criação do cluster.

Se você estiver usando uma EMR versão mais antiga da Amazon que não tem a solução permanente para esse problema, a solução alternativa a seguir permite que você defina explicitamente o ulimit do controlador de instância para um máximo de 65536 arquivos.

Defina explicitamente um ulimit na linha de comando

1. Edite `/etc/systemd/system/instance-controller.service` para adicionar os seguintes parâmetros à seção Serviço.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Reiniciar InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Defina um ulimit usando a ação de bootstrap (BA)

Você também pode usar um script de ação de bootstrap (BA) para configurar o ulimit instance-controller para 65536 arquivos na criação do 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

EMRclusters que executam Amazon Linux ou Amazon Linux 2 Amazon Machine Images (AMIs) usam o comportamento padrão do Amazon Linux e não baixam e instalam automaticamente atualizações importantes e críticas do kernel que exigem uma reinicialização. Esse é o mesmo comportamento de outras EC2 instâncias da Amazon que executam o Amazon Linux padrãoAMI. Se novas atualizações de software Amazon Linux que exigem uma reinicialização (como kernel e CUDA atualizações) ficarem disponíveis após a disponibilização de uma EMR versão da Amazon, as instâncias de EMR cluster que executam o padrão AMI não baixam e instalam automaticamente essas atualizações. NVIDIA Para obter atualizações do kernel, você pode [personalizar sua Amazon EMR AMI](#) para [usar o Amazon Linux AMI mais recente](#).

- O suporte do console para criar uma configuração de segurança que especifica a opção de integração do AWS Ranger atualmente não é suportado na GovCloud região. A configuração de segurança pode ser feita usando CLI o. Consulte [Criar a configuração EMR de segurança](#) no Guia EMR de gerenciamento da Amazon.
- Quando AtRestEncryption ou a HDFS criptografia é habilitada em um cluster que usa o Amazon EMR 5.31.0 ou 5.32.0, as consultas do Hive resultam na seguinte exceção de tempo de execução.

```
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
```

- Quando você usa o Spark com a formatação de localização de partições do Hive para ler dados no Amazon S3 e executa o Spark nas EMR versões 5.30.0 a 5.36.0 e 6.2.0 a 6.9.0 da Amazon, você pode encontrar um problema que impede seu cluster de ler os dados corretamente. Isso poderá acontecer se suas partições tiverem todas as características a seguir:
 - Duas ou mais partições são verificadas na mesma tabela.

- Pelo menos um caminho de diretório de partição é um prefixo de pelo menos outro caminho de diretório de partição, por exemplo, `s3://bucket/table/p=a` é um prefixo de `s3://bucket/table/p=a b`.
- O primeiro caractere que segue o prefixo no outro diretório de partição tem um valor UTF -8 menor que o / caractere (U+002F). Por exemplo, o caractere de espaço (U+0020) que ocorre entre a e b em `s3://bucket/table/p=a b` se enquadra nessa categoria. Observe que existem 14 outros caracteres que não são de controle: `!"#$%&'()*+,-.` Para obter mais informações, consulte a [tabela de codificação UTF -8 e caracteres Unicode](#).

Como solução alternativa para esse problema, defina a configuração `spark.sql.sources.fastS3PartitionDiscovery.enabled` como `false` na classificação `spark-defaults`.

Versões de componente 5.32.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.4.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.16.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
emr-goodies	2.13.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.5.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-notebook-env	1.1.0	Ambiente Conda para bloco de anotações do emr, que inclui o jupyter enterprise gateway
emr-s3-dist-cp	2.17.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	1.6.0	EMRConector S3Select
emrfs	2.45.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.11.2	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
flink-jobmanager-config	1.11.2	Gerenciando recursos em EMR nós para o Apache JobManager Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.

Componente	Version (Versão)	Descrição
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.10.1-amzn-0	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.10.1-amzn-0	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.10.1-amzn-0	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.10.1-amzn-0	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	2.10.1-amzn-0	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httpfs-server	2.10.1-amzn-0	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.10.1-amzn-0	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API

Componente	Version (Versão)	Descrição
hadoop-mapred	2.10.1-amzn-0	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.10.1-amzn-0	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.10.1-amzn-0	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.10.1-amzn-0	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.13	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.13	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.4.13	HBasecliente de linha de comando.
hbase-rest-server	1.4.13	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	1.4.13	Serviço que fornece um endpoint Thrift para. HBase

Componente	Version (Versão)	Descrição
hcatalog-client	2.3.7-amzn-3	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.7-amzn-3	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.7-amzn-3	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	2.3.7-amzn-3	O cliente da linha de comando do Hive.
hive-hbase	2.3.7-amzn-3	Cliente do Hive-hbase.
hive-metastore-server	2.3.7-amzn-3	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.7-amzn-3	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.6.0-amzn-0	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.
hudi-spark	0.6.0-amzn-0	Biblioteca de pacotes para executar o Spark com o Hudi.

Componente	Version (Versão)	Descrição
hudi-presto	0.6.0-amzn-0	Biblioteca de pacotes para executar o Presto com o Hudi.
hue-server	4.8.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	1.1.0	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.7.0-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.7.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mariadb-server	5.5.68	Meu servidor SQL de banco de dados.
nvidia-cuda	10.1.243	Drivers NVIDIA e toolkit CUDA
oozie-client	5.2.0	O cliente da linha de comando do Oozie.
oozie-server	5.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.

Componente	Version (Versão)	Descrição
opencv	4.4.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.3- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.3- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.240.1-amzn-0	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.240.1-amzn-0	O serviço que executa partes de uma consulta.
presto-client	0.240.1-amzn-0	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.3	O projeto R para computação estatística
ranger-kms-server	1.2.0	Sistema de gerenciamento de chaves Apache Ranger
spark-client	2.4.7-amzn-0	Os clientes da linha de comando do Spark.

Componente	Version (Versão)	Descrição
spark-history-server	2.4.7-amzn-0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.4.7-amzn-0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.4.7-amzn-0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	2.3.1	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.2	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.
zeppelin-server	0.8.2	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.14	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.

Componente	Version (Versão)	Descrição
zookeeper-client	3.4.14	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.32.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Ocorrem ações de reconfiguração quando você especifica uma configuração para grupos de instâncias em um cluster em execução. A Amazon EMR só inicia ações de reconfiguração para as classificações que você modifica. Para ter mais informações, consulte [Reconfigurar um grupo de instâncias de um cluster em execução](#).

Classificações do emr-5.32.0

Classificações	Descrição	Ações de reconfiguração
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.	Restarts the ResourceManager service.
container-executor	Altere os valores no arquivo <code>container-executor.cfg</code> YARN do Hadoop.	Not available.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.	Not available.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Restarts the Hadoop YARN services ResourceManager,

Classificações	Descrição	Ações de reconfiguração
		NodeManager, ProxyServer, and TimelineServer. Additionally restarts Hadoop KMS, Ranger KMS, HiveServer2, Hive MetaStore, Hadoop Httpfs, and MapReduce-HistoryServer.
docker-conf	Alterar as configurações relacionadas ao docker.	Not available.
emrfs-site	Altere EMRFS as configurações.	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	Alterar as configurações do flink-conf.yaml.	Restarts Flink history server.
flink-log4j	Alterar as configurações de log4j.properties no Flink.	Restarts Flink history server.
flink-log4j-yarn-session	Altere as configurações do Flink log4j-yarn-session.properties.	Not available.

Classificações	Descrição	Ações de reconfiguração
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.	Restarts Flink history server.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do 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	Alterar os valores no arquivo log4j.properties do 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	Alterar a configuração do servidor SSL no Hadoop	Not available.
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop	Not available.

Classificações	Descrição	Ações de reconfiguração
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase	Custom EMR specific property. Sets emrfs-site and hbase-site configs. See those for their associated restarts.
hbase-env	Mude os valores no HBase ambiente.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.	Not available.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.	Restarts the HBase services RegionServer, HBaseMaster, ThriftServer, RestServer. Additionally restarts Phoenix QueryServer.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.	Should not be reconfigured.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.	Restarts the Hadoop HDFS services Namenode, SecondaryNamenode, Datanode, ZKFC, and Journalnode. Additionally restarts Hadoop Httpfs.

Classificações	Descrição	Ações de reconfiguração
hcatalog-env	Mude os valores no HCatalog ambiente.	Restarts Hive HCatalog Server.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.	Restarts Hive HCatalog Server.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.	Restarts Hive HCatalog Server.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.	Restarts Hive WebHCat Server.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat	Restarts Hive WebHCat Server.
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.	Not available.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.	Not available.
hive-env	Alterar os valores no ambiente do Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log 4j2.properties do Hive.	Restarts HiveServer2 and HiveMetastore.

Classificações	Descrição	Ações de reconfiguração
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log 4j2.properties do Hive.	Not available.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.	Not available.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.	Restarts HiveServer2 and HiveMetastore. Runs Hive schemaTool CLI commands to verify hive-metastore. Also restarts Oozie and Zeppelin.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.	Not available.
hue-ini	Alterar os valores no arquivo ini do Hue	Restarts Hue. Also activates Hue config override CLI commands to pick up new configurations.
httpfs-env	Mude os valores no HTTPFS ambiente.	Restarts Hadoop Httpfs service.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.	Restarts Hadoop Httpfs service.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.	Not available.
hadoop-kms-env	Altere os valores no ambiente do HadoopKMS.	Restarts Hadoop-KMS service.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.	Not available.

Classificações	Descrição	Ações de reconfiguração
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.	Restarts Hadoop-KMS and Ranger-KMS service.
hudi-env	Altere os valores no ambiente do Hudi.	Not available.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.	Not available.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.	Not available.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.	Not available.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.	Not available.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.	Restarts Livy Server.
livy-env	Alterar os valores no ambiente do Livy.	Restarts Livy Server.
livy-log4j	Alterar as configurações de log4j.properties no Livy.	Restarts Livy Server.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.	Restarts Hadoop MapReduce-HistoryServer.

Classificações	Descrição	Ações de reconfiguração
oozie-env	Alterar os valores no ambiente do Oozie.	Restarts Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.	Restarts Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.	Restarts Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.	Not available.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.	Not available.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.	Restarts Phoenix-QueryServer.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.	Not available.
pig-env	Alterar os valores no ambiente do Pig.	Not available.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.	Restarts Oozie.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.	Not available.
presto-log	Alterar os valores no arquivo log.properties do Presto.	Restarts Presto-Server.

Classificações	Descrição	Ações de reconfiguração
presto-config	Alterar os valores no arquivo <code>config.properties</code> do Presto.	Restarts Presto-Server.
presto-password-authenticator	Altere os valores no arquivo <code>password-authenticator.properties</code> do Presto.	Not available.
presto-env	Altere os valores no arquivo <code>presto-env.sh</code> do Presto.	Restarts Presto-Server.
presto-node	Altere os valores no arquivo <code>node.properties</code> do Presto.	Not available.
presto-connector-blackhole	Alterar os valores no arquivo <code>blackhole.properties</code> do Presto.	Not available.
presto-connector-cassandra	Alterar os valores no arquivo <code>cassandra.properties</code> do Presto.	Not available.
presto-connector-hive	Alterar os valores no arquivo <code>hive.properties</code> do Presto.	Restarts Presto-Server.
presto-connector-jmx	Alterar os valores no arquivo <code>jmx.properties</code> do Presto.	Not available.
presto-connector-kafka	Alterar os valores no arquivo <code>kafka.properties</code> do Presto.	Not available.
presto-connector-localfile	Alterar os valores no arquivo <code>localfile.properties</code> do Presto.	Not available.
presto-connector-memory	Alterar os valores no arquivo <code>memory.properties</code> do Presto.	Not available.

Classificações	Descrição	Ações de reconfiguração
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.	Not available.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.	Not available.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.	Not available.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.	Not available.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.	Not available.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.	Not available.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.	Not available.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.	Not available.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS	Restarts Ranger KMS Server.
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.	Restarts Ranger KMS Server.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.	Restarts Ranger KMS Server.

Classificações	Descrição	Ações de reconfiguração
ranger-kms-log4j	Altere os valores no arquivo kms-log4j.properties do Ranger. KMS	Not available.
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.	Not available.
recordserver-env	Mude os valores no EMR RecordServer ambiente.	Restarts EMR record server.
recordserver-conf	Altere os valores no arquivo EMR RecordServer server.properties do.	Restarts EMR record server.
recordserver-log4j	Altere os valores no EMR RecordServer arquivo log4j.properties.	Restarts EMR record server.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.	This property modifies spark-defaults. See actions there.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.	Restarts Spark history server and Spark thrift server.
spark-env	Alterar os valores no ambiente do Spark.	Restarts Spark history server and Spark thrift server.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.	Not available.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.	Restarts Spark history server and Spark thrift server.

Classificações	Descrição	Ações de reconfiguração
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.	Restarts Spark history server and Spark thrift server.
sqoop-env	Alterar os valores no ambiente do Sqoop.	Not available.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.	Not available.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.	Not available.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.	Restarts Oozie.
yarn-env	Mude os valores no YARN ambiente.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts MapReduce-HistoryServer.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.	Restarts the Hadoop YARN services ResourceManager, NodeManager, ProxyServer, and TimelineServer. Additionally restarts Livy Server and MapReduce-HistoryServer.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.	Restarts Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.	Restarts Zookeeper server.

Classificações	Descrição	Ações de reconfiguração
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.	Restarts Zookeeper server.

Amazon EMR versão 5.31.1

Versões 5.31.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.31.1	emr-5.31.0	emr-5.30.2	emr-5.30.1
AWS SDK para Java	1.11.852	1.11.852	1.11.759	1.11.759
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.11.12	2.11.12	2.11.12	2.11.12
AmazonCloudWatchAgent	-	-	-	-

	emr-5.31.1	emr-5.31.0	emr-5.30.2	emr-5.30.1
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
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

	emr-5.31.1	emr-5.31.0	emr-5.30.2	emr-5.30.1
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
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

Notas da versão 5.31.1

Esta é uma versão para corrigir problemas com o Amazon EMR Scaling quando ele falha na escalabilidade ou redução de escala de um cluster com sucesso ou causa falhas no aplicativo.

Alterações, melhorias e problemas resolvidos

- Corrigido um problema em que as solicitações de escalabilidade falhavam em um cluster grande e altamente utilizado quando os daemons no cluster da EMR Amazon estavam executando atividades de verificação de integridade, como coletar o estado do nó e o estado do YARN nó. HDFS Isso estava acontecendo porque os daemons no cluster não conseguiam comunicar os dados do status de saúde de um nó aos componentes internos da Amazon. EMR
- Daemons EMR no cluster aprimorados para rastrear corretamente os estados dos nós quando os endereços IP são reutilizados para melhorar a confiabilidade durante as operações de escalabilidade.
- [SPARK-29683](#). Corrigido um problema em que ocorriam falhas de trabalho durante a redução da escala verticalmente do cluster, pois o Spark presumia que todos os nós disponíveis estavam na lista de negação.
- [YARN-9011](#). Corrigido um problema em que falhas de trabalho ocorriam devido a uma condição de corrida no YARN descomissionamento quando o cluster tentava aumentar ou diminuir a escala.
- Foi corrigido o problema com falhas de etapas ou tarefas durante a escalabilidade do cluster, garantindo que os estados dos nós fossem sempre consistentes entre os EMR daemons no cluster da Amazon e/. YARN HDFS

- Corrigido um problema em que operações de cluster, como redução de escala e envio de etapas, falhavam em EMR clusters da Amazon habilitados com a autenticação Kerberos. Isso ocorreu porque o daemon no EMR cluster da Amazon não renovou o tíquete Kerberos, que é necessário para se comunicar de forma segura com/em execução no nó primário. HDFS YARN
- As EMR versões mais recentes da Amazon corrigem o problema com um limite menor de “Máximo de arquivos abertos” em relação aos mais antigos AL2 na AmazonEMR. As EMR versões 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 e posteriores agora incluem uma correção permanente com uma configuração mais alta de “Máximo de arquivos abertos”.
- HTTPS agora está habilitado por padrão para repositórios Amazon Linux. Se você estiver usando uma VPCE política do Amazon S3 para restringir o acesso a buckets específicos, deverá adicionar o novo bucket Amazon Linux ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` à sua política (`$region` substitua pela região onde está o endpoint). Para obter mais informações, consulte esse tópico nos fóruns de AWS discussão. [Anúncio: o Amazon Linux 2 agora oferece suporte à capacidade de uso HTTPS ao se conectar a repositórios de pacotes.](#)

Problemas conhecidos

- Quando você usa o Spark com a formatação de localização de partições do Hive para ler dados no Amazon S3 e executa o Spark nas EMR versões 5.30.0 a 5.36.0 e 6.2.0 a 6.9.0 da Amazon, você pode encontrar um problema que impede seu cluster de ler os dados corretamente. Isso poderá acontecer se suas partições tiverem todas as características a seguir:
 - Duas ou mais partições são verificadas na mesma tabela.
 - Pelo menos um caminho de diretório de partição é um prefixo de pelo menos outro caminho de diretório de partição, por exemplo, `s3://bucket/table/p=a` é um prefixo de `s3://bucket/table/p=a b`.
 - O primeiro caractere que segue o prefixo no outro diretório de partição tem um valor UTF -8 menor que o / caractere (U+002F). Por exemplo, o caractere de espaço (U+0020) que ocorre entre a e b em `s3://bucket/table/p=a b` se enquadra nessa categoria. Observe que existem 14 outros caracteres que não são de controle: `!"#$%&'()*+,-.` Para obter mais informações, consulte a [tabela de codificação UTF -8 e caracteres Unicode](#).

Como solução alternativa para esse problema, defina a configuração `spark.sql.sources.fastS3PartitionDiscovery.enabled` como `false` na classificação `spark-defaults`.

Versões de componente 5.31.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.4.0	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.15.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.13.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.5.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.15.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
<code>emr-s3-select</code>	1.6.0	EMRConector S3Select
<code>emrfs</code>	2.43.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
flink-client	1.11.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
flink-jobmanager-config	1.11.0	Gerenciando recursos em EMR nós para o Apache JobManager Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.10.0-amzn-0.1	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.10.0-amzn-0.1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.10.0-amzn-0.1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.10.0-amzn-0.1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.

Componente	Version (Versão)	Descrição
hadoop-hdfs-journalnode	2.10.0-amzn-0.1	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httfs-server	2.10.0-amzn-0.1	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.10.0-amzn-0.1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.10.0-amzn-0.1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.10.0-amzn-0.1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.10.0-amzn-0.1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.10.0-amzn-0.1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.13	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.13	Serviço para atender a uma ou mais HBase regiões.

Componente	Version (Versão)	Descrição
hbase-client	1.4.13	HBase cliente de linha de comando.
hbase-rest-server	1.4.13	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.4.13	Serviço que fornece um endpoint Thrift para HBase
hcatalog-client	2.3.7-amzn-1	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.7-amzn-1	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.7-amzn-1	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.3.7-amzn-1	O cliente da linha de comando do Hive.
hive-hbase	2.3.7-amzn-1	Cliente do Hive-hbase.
hive-metastore-server	2.3.7-amzn-1	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL

Componente	Version (Versão)	Descrição
hive-server2	2.3.7-amzn-1	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.6.0-amzn-0	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.
hudi-spark	0.6.0-amzn-0	Biblioteca de pacotes para executar o Spark com o Hudi.
hudi-presto	0.6.0-amzn-0	Biblioteca de pacotes para executar o Presto com o Hudi.
hue-server	4.7.1	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	1.1.0	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.7.0-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.6.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.

Componente	Version (Versão)	Descrição
mariadb-server	5.5.64+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.2.88	Drivers NVIDIA e toolkit CUDA
oozie-client	5.2.0	O cliente da linha de comando do Oozie.
oozie-server	5.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	4.3.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.3- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.3- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.238.3-amzn-0	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.238.3-amzn-0	O serviço que executa partes de uma consulta.
presto-client	0.238.3-amzn-0	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.

Componente	Version (Versão)	Descrição
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.3	O projeto R para computação estatística
ranger-kms-server	1.2.0	Sistema de gerenciamento de chaves Apache Ranger
spark-client	2.4.6-amzn-0.1	Os clientes da linha de comando do Spark.
spark-history-server	2.4.6-amzn-0.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.4.6-amzn-0.1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.4.6-amzn-0.1	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	2.1.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.2	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.

Componente	Version (Versão)	Descrição
zeppelin-server	0.8.2	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.14	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.14	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.31.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.31.1

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .

Classificações	Descrição
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.

Classificações	Descrição
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.

Classificações	Descrição
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente do HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
hudi-env	Altere os valores no ambiente do Hudi.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.

Classificações	Descrição
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.

Classificações	Descrição
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.

Classificações	Descrição
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.
ranger-kms-log4j	Altere os valores no arquivo kms-log4j.properties do Ranger. KMS
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.

Classificações	Descrição
recordserver-env	Mude os valores no EMR RecordServer ambiente.
recordserver-conf	Altere os valores no arquivo EMR RecordServer <code>erver.properties</code> do.
recordserver-log4j	Altere os valores no EMR RecordServer arquivo <code>log4j.properties</code> .
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo <code>spark-defaults.conf</code> do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo <code>hive-site.xml</code> do Spark.
spark-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Spark.
spark-metrics	Alterar os valores no arquivo <code>metrics.properties</code> do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo <code>oraoop-site.xml</code> OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo <code>sqoop-site.xml</code> do Sqoop.
tez-site	Alterar os valores no arquivo <code>tez-site.xml</code> do Tez.
yarn-env	Mude os valores no YARN ambiente.

Classificações	Descrição
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.31.0

Versões 5.31.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.31.0	emr-5.30.2	emr-5.30.1	emr-5.30.0
AWS SDK para Java	1.11.852	1.11.759	1.11.759	1.11.759

	emr-5.31.0	emr-5.30.2	emr-5.30.1	emr-5.30.0
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,7
Scala	2.11.12	2.11.12	2.11.12	2.11.12
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
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

	emr-5.31.0	emr-5.30.2	emr-5.30.1	emr-5.30.0
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

Notas da versão 5.31.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.31.0 da Amazon. As alterações são referentes à versão 5.30.1.

Data da versão inicial: 9 de outubro de 2020

Data da última atualização: 15 de outubro de 2020

Atualizações

- Atualizado o conector do Amazon Glue para a versão 1.13.0
- Amazon SageMaker Spark atualizado SDK para a versão 1.4.0
- Atualizado o conector do Amazon Kinesis para a versão 3.5.9
- Atualizado para AWS SDK for Java a versão 1.11.852
- Atualizado o Bigtop-tomcat para a versão 8.5.56

- EMRFS atualizado para a versão 2.43.0
- EMR MetricsAndEventsApiGateway Cliente atualizado para a versão 1.4.0
- EMRS3 Dist CP atualizado para a versão 2.15.0
- EMRS3 Select atualizado para a versão 1.6.0
- Atualizado o Flink para a versão 1.11.0
- Atualizado o Hadoop para a versão 2.10.0
- Atualizado o Hive para a versão 2.3.7
- Atualizado o Hudi para a versão 0.6.0
- Atualizado o Hue para a versão 4.7.1
- Atualizado JupyterHub para a versão 1.1.0
- Atualizado o Mxnet para a versão 1.6.0
- Atualizado o OpenCV para a versão 4.3.0
- Atualizado o Presto para a versão 0.238.3
- Atualizado TensorFlow para a versão 2.1.0

Alterações, melhorias e problemas resolvidos

- Esta é uma versão para corrigir problemas com o Amazon EMR Scaling quando ele falha na escalabilidade ou redução de escala de um cluster com sucesso ou causa falhas no aplicativo.
- Corrigido um problema em que as solicitações de escalabilidade falhavam em um cluster grande e altamente utilizado quando os daemons no cluster da EMR Amazon estavam executando atividades de verificação de integridade, como coletar o estado do nó e o estado do YARN nó. HDFS Isso estava acontecendo porque os daemons no cluster não conseguiam comunicar os dados do status de saúde de um nó aos componentes internos da Amazon. EMR
- Daemons EMR no cluster aprimorados para rastrear corretamente os estados dos nós quando os endereços IP são reutilizados para melhorar a confiabilidade durante as operações de escalabilidade.
- [SPARK-29683](#). Corrigido um problema em que ocorriam falhas de trabalho durante a redução da escala verticalmente do cluster, pois o Spark presumia que todos os nós disponíveis estavam na lista de negação.
- [YARN-9011](#). Corrigido um problema em que falhas de trabalho ocorriam devido a uma condição de corrida no YARN descomissionamento quando o cluster tentava aumentar ou diminuir a escala.

- Foi corrigido o problema com falhas de etapas ou tarefas durante a escalabilidade do cluster, garantindo que os estados dos nós fossem sempre consistentes entre os EMR daemons no cluster da Amazon e/. YARN HDFS
- Corrigido um problema em que operações de cluster, como redução de escala e envio de etapas, falhavam em EMR clusters da Amazon habilitados com a autenticação Kerberos. Isso ocorreu porque o daemon no EMR cluster da Amazon não renovou o tíquete Kerberos, que é necessário para se comunicar de forma segura com/em execução no nó primário. HDFS YARN
- As EMR versões mais recentes da Amazon corrigem o problema com um limite menor de “Máximo de arquivos abertos” em relação aos mais antigos AL2 na AmazonEMR. As EMR versões 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 e posteriores agora incluem uma correção permanente com uma configuração mais alta de “Máximo de arquivos abertos”.
- [As estatísticas de colunas do Hive](#) são compatíveis com as EMR versões 5.31.0 e posteriores da Amazon.
- Atualizadas versões do componente.
- EMRFSSuporte ao S3EC V2 na Amazon EMR 5.31.0. Nas SDK versões 1.11.837 e posteriores do Java do S3, a versão 2 do cliente de criptografia (S3EC V2) foi introduzida com vários aprimoramentos de segurança. Para obter mais informações, consulte as informações a seguir.
 - Publicações no blog do S3: [Atualizações no cliente de criptografia do Amazon S3](#).
 - AWS SDK for Java Guia do desenvolvedor: [Migre clientes de criptografia e descriptografia](#) para a V2.
 - EMRGuia de gerenciamento: criptografia [do lado do cliente do Amazon S3](#).

O Encryption Client V1 ainda está disponível no SDK para compatibilidade com versões anteriores.

Novos atributos

- Limite inferior de “Máximo de arquivos abertos” em versões mais antigas AL2 [corrigido em versões mais recentes]. EMRLançamentos da Amazon: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 e emr-6.2.0 são baseados em versões mais antigas do Linux ofAmazon 2 ()AL2, que têm uma configuração de limite inferior para “Máximo de arquivos abertos” quando os clusters da Amazon são criados com o padrão. EMR AMI As EMR versões 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 e posteriores incluem uma correção permanente com uma configuração maior de “Máximo de arquivos abertos”. Versões com o limite inferior de arquivos abertos causam o erro “Muitos arquivos abertos” ao ser enviado um trabalho do Spark. Nas versões afetadas, o padrão da Amazon AMI tem uma configuração EMR padrão de ulimit

de 4096 para “Máximo de arquivos abertos”, que é inferior ao limite de 65536 arquivos no Linux 2. latestAmazon AMI A configuração inferior de ulimit para “Máximo de arquivos abertos” causa falhas em trabalhos do Spark quando o driver e o executor do Spark tentam abrir mais de 4096 arquivos. Para corrigir o problema, a Amazon EMR tem um script de ação de bootstrap (BA) que ajusta a configuração ulimit na criação do cluster.

Se você estiver usando uma EMR versão mais antiga da Amazon que não tem a solução permanente para esse problema, a solução alternativa a seguir permite que você defina explicitamente o ulimit do controlador de instância para um máximo de 65536 arquivos.

Defina explicitamente um ulimit na linha de comando

1. Edite `/etc/systemd/system/instance-controller.service` para adicionar os seguintes parâmetros à seção Serviço.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Reiniciar InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Defina um ulimit usando a ação de bootstrap (BA)

Você também pode usar um script de ação de bootstrap (BA) para configurar o ulimit instance-controller para 65536 arquivos na criação do 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
```

- Com o Amazon EMR 5.31.0, você pode lançar um cluster que se integra ao Lake Formation. Essa integração fornece filtragem de dados refinada em nível de coluna para bancos de dados e tabelas no Glue Data Catalog. AWS Ele também permite o login único federado em EMR notebooks ou Apache Zeppelin a partir de um sistema de identidade corporativo. Para obter mais informações, consulte [Integração da Amazon EMR com](#) o Guia AWS Lake Formation de EMR gerenciamento da Amazon.

Atualmente, o Amazon EMR with Lake Formation está disponível em 16 AWS regiões: Leste dos EUA (Ohio e Norte da Virgínia), Oeste dos EUA (Norte da Califórnia e Oregon), Ásia-Pacífico (Mumbai, Seul, Cingapura, Sydney e Tóquio), Canadá (Central), Europa (Frankfurt, Irlanda, Londres, Paris e Estocolmo) e América do Sul (São Paulo).

Problemas conhecidos

- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSH como hadoop usuário do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do hadoop.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo keytab está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

- Quando `AtRestEncryption` ou a HDFS criptografia é habilitada em um cluster que usa o Amazon EMR 5.31.0 ou 5.32.0, as consultas do Hive resultam na seguinte exceção de tempo de execução.

```
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
```

- Quando você usa o Spark com a formatação de localização de partições do Hive para ler dados no Amazon S3 e executa o Spark nas EMR versões 5.30.0 a 5.36.0 e 6.2.0 a 6.9.0 da Amazon, você pode encontrar um problema que impede seu cluster de ler os dados corretamente. Isso poderá acontecer se suas partições tiverem todas as características a seguir:
 - Duas ou mais partições são verificadas na mesma tabela.
 - Pelo menos um caminho de diretório de partição é um prefixo de pelo menos outro caminho de diretório de partição, por exemplo, `s3://bucket/table/p=a` é um prefixo de `s3://bucket/table/p=a b`.
 - O primeiro caractere que segue o prefixo no outro diretório de partição tem um valor UTF -8 menor que o `/` caractere (U+002F). Por exemplo, o caractere de espaço (U+0020) que ocorre entre `a` e `b` em `s3://bucket/table/p=a b` se enquadra nessa categoria. Observe que existem 14 outros caracteres que não são de controle: `!"#$%&'()*+,-.` Para obter mais informações, consulte a [tabela de codificação UTF -8 e caracteres Unicode](#).

Como solução alternativa para esse problema, defina a configuração `spark.sql.sources.fastS3PartitionDiscovery.enabled` como `false` na classificação `spark-defaults`.

Versões de componente 5.31.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion-amzn-EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.4.0	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.15.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.13.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.5.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.15.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
<code>emr-s3-select</code>	1.6.0	EMRConector S3Select
<code>emrfs</code>	2.43.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
flink-client	1.11.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
flink-jobmanager-config	1.11.0	Gerenciando recursos em EMR nós para o Apache JobManager Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.10.0-amzn-0	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.10.0-amzn-0	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.10.0-amzn-0	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.10.0-amzn-0	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.

Componente	Version (Versão)	Descrição
hadoop-hdfs-journalnode	2.10.0-amzn-0	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httfs-server	2.10.0-amzn-0	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.10.0-amzn-0	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.10.0-amzn-0	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.10.0-amzn-0	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.10.0-amzn-0	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.10.0-amzn-0	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.13	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.13	Serviço para atender a uma ou mais HBase regiões.

Componente	Version (Versão)	Descrição
hbase-client	1.4.13	HBase cliente de linha de comando.
hbase-rest-server	1.4.13	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.4.13	Serviço que fornece um endpoint Thrift para HBase
hcatalog-client	2.3.7-amzn-1	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.7-amzn-1	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.7-amzn-1	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.3.7-amzn-1	O cliente da linha de comando do Hive.
hive-hbase	2.3.7-amzn-1	Cliente do Hive-hbase.
hive-metastore-server	2.3.7-amzn-1	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL

Componente	Version (Versão)	Descrição
hive-server2	2.3.7-amzn-1	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.6.0-amzn-0	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.
hudi-spark	0.6.0-amzn-0	Biblioteca de pacotes para executar o Spark com o Hudi.
hudi-presto	0.6.0-amzn-0	Biblioteca de pacotes para executar o Presto com o Hudi.
hue-server	4.7.1	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	1.1.0	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.7.0-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.6.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.

Componente	Version (Versão)	Descrição
mariadb-server	5.5.64	Meu servidor SQL de banco de dados.
nvidia-cuda	9.2.88	Drivers NVIDIA e toolkit CUDA
oozie-client	5.2.0	O cliente da linha de comando do Oozie.
oozie-server	5.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	4.3.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.3- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.3- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.238.3-amzn-0	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.238.3-amzn-0	O serviço que executa partes de uma consulta.
presto-client	0.238.3-amzn-0	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.

Componente	Version (Versão)	Descrição
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.3	O projeto R para computação estatística
ranger-kms-server	1.2.0	Sistema de gerenciamento de chaves Apache Ranger
spark-client	2.4.6-amzn-0	Os clientes da linha de comando do Spark.
spark-history-server	2.4.6-amzn-0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.4.6-amzn-0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.4.6-amzn-0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	2.1.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.2	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.

Componente	Version (Versão)	Descrição
zeppelin-server	0.8.2	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.14	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.14	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.31.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.31.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .

Classificações	Descrição
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.

Classificações	Descrição
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.

Classificações	Descrição
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente do HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
hudi-env	Altere os valores no ambiente do Hudi.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.

Classificações	Descrição
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.

Classificações	Descrição
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.

Classificações	Descrição
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.
ranger-kms-log4j	Altere os valores no arquivo kms-log4j.properties do Ranger. KMS
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.

Classificações	Descrição
recordserver-env	Mude os valores no EMR RecordServer ambiente.
recordserver-conf	Altere os valores no arquivo EMR RecordServer <code>erver.properties</code> do.
recordserver-log4j	Altere os valores no EMR RecordServer arquivo <code>log4j.properties</code> .
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo <code>spark-defaults.conf</code> do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo <code>hive-site.xml</code> do Spark.
spark-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Spark.
spark-metrics	Alterar os valores no arquivo <code>metrics.properties</code> do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo <code>oraoop-site.xml</code> OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo <code>sqoop-site.xml</code> do Sqoop.
tez-site	Alterar os valores no arquivo <code>tez-site.xml</code> do Tez.
yarn-env	Mude os valores no YARN ambiente.

Classificações	Descrição
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.30.2

Versões 5.30.2 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.30.2	emr-5.30.1	emr-5.30.0	emr-5.29.0
AWS SDK para Java	1.11.759	1.11.759	1.11.759	1.11.682

	emr-5.30.2	emr-5.30.1	emr-5.30.0	emr-5.29.0
Python	2,7, 3,7	2,7, 3,7	2,7, 3,7	2,7, 3,6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
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
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

	emr-5.30.2	emr-5.30.1	emr-5.30.0	emr-5.29.0
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

Notas da versão 5.30.2

Esta é uma versão para corrigir problemas com o Amazon EMR Scaling quando ele falha na escalabilidade ou redução de escala de um cluster com sucesso ou causa falhas no aplicativo.

Alterações, melhorias e problemas resolvidos

- Corrigido um problema em que as solicitações de escalabilidade falhavam em um cluster grande e altamente utilizado quando os daemons no cluster da EMR Amazon estavam executando atividades de verificação de integridade, como coletar o estado do nó e o estado do YARN nó. HDFS Isso estava acontecendo porque os daemons no cluster não conseguiam comunicar os dados do status de saúde de um nó aos componentes internos da Amazon. EMR
- Daemons EMR no cluster aprimorados para rastrear corretamente os estados dos nós quando os endereços IP são reutilizados para melhorar a confiabilidade durante as operações de escalabilidade.

- [SPARK-29683](#). Corrigido um problema em que ocorriam falhas de trabalho durante a redução da escala verticalmente do cluster, pois o Spark presumia que todos os nós disponíveis estavam na lista de negação.
- [YARN-9011](#). Corrigido um problema em que falhas de trabalho ocorriam devido a uma condição de corrida no YARN descomissionamento quando o cluster tentava aumentar ou diminuir a escala.
- Foi corrigido o problema com falhas de etapas ou tarefas durante a escalabilidade do cluster, garantindo que os estados dos nós fossem sempre consistentes entre os EMR daemons no cluster da Amazon e/. YARN HDFS
- Corrigido um problema em que operações de cluster, como redução de escala e envio de etapas, falhavam em EMR clusters da Amazon habilitados com a autenticação Kerberos. Isso ocorreu porque o daemon no EMR cluster da Amazon não renovou o tíquete Kerberos, que é necessário para se comunicar de forma segura com/em execução no nó primário. HDFS YARN
- As EMR versões mais recentes da Amazon corrigem o problema com um limite menor de “Máximo de arquivos abertos” em relação aos mais antigos AL2 na AmazonEMR. As EMR versões 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 e posteriores agora incluem uma correção permanente com uma configuração mais alta de “Máximo de arquivos abertos”.
- HTTPS agora está habilitado por padrão para repositórios Amazon Linux. Se você estiver usando uma VPCE política do Amazon S3 para restringir o acesso a buckets específicos, deverá adicionar o novo bucket Amazon Linux ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` à sua política (`$region` substitua pela região onde está o endpoint). Para obter mais informações, consulte esse tópico nos fóruns de AWS discussão. [Anúncio: o Amazon Linux 2 agora oferece suporte à capacidade de uso HTTPS ao se conectar a repositórios de pacotes.](#)

Problemas conhecidos

- Quando você usa o Spark com a formatação de localização de partições do Hive para ler dados no Amazon S3 e executa o Spark nas EMR versões 5.30.0 a 5.36.0 e 6.2.0 a 6.9.0 da Amazon, você pode encontrar um problema que impede seu cluster de ler os dados corretamente. Isso poderá acontecer se suas partições tiverem todas as características a seguir:
 - Duas ou mais partições são verificadas na mesma tabela.
 - Pelo menos um caminho de diretório de partição é um prefixo de pelo menos outro caminho de diretório de partição, por exemplo, `s3://bucket/table/p=a` é um prefixo de `s3://bucket/table/p=a b`.
 - O primeiro caractere que segue o prefixo no outro diretório de partição tem um valor UTF-8 menor que o / caractere (U+002F). Por exemplo, o caractere de espaço (U+0020) que ocorre

entre a e b em `s3://bucket/table/p=a b` se enquadra nessa categoria. Observe que existem 14 outros caracteres que não são de controle: `!"#$%&'()*+,-.` Para obter mais informações, consulte a [tabela de codificação UTF -8 e caracteres Unicode](#).

Como solução alternativa para esse problema, defina a configuração `spark.sql.sources.fastS3PartitionDiscovery.enabled` como `false` na classificação `spark-defaults`.

Versões de componente 5.30.2

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.3.0	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.14.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.13.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.5.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
emr-s3-dist-cp	2.14.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	1.5.0	EMRConector S3Select
emrfs	2.40.0	O conector do Amazon S3 para aplicativos do ecossistema do Hadoop.
flink-client	1.10.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.5-amzn-6.1	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-6.1	HDFSserviço em nível de nó para armazenar blocos.

Componente	Version (Versão)	Descrição
hadoop-hdfs-library	2.8.5-amzn-6.1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.8.5-amzn-6.1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	2.8.5-amzn-6.1	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-https-server	2.8.5-amzn-6.1	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.8.5-amzn-6.1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-6.1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.8.5-amzn-6.1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.8.5-amzn-6.1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.5-amzn-6.1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.

Componente	Version (Versão)	Descrição
hbase-hmaster	1.4.13	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.13	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.4.13	HBase cliente de linha de comando.
hbase-rest-server	1.4.13	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.4.13	Serviço que fornece um endpoint Thrift para HBase.
hcatalog-client	2.3.6-amzn-2	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.6-amzn-2	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.6-amzn-2	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.3.6-amzn-2	O cliente da linha de comando do Hive.
hive-hbase	2.3.6-amzn-2	Cliente do Hive-hbase.

Componente	Version (Versão)	Descrição
hive-metastore-server	2.3.6-amzn-2	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.6-amzn-2	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.5.2-incubating	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.
hudi-presto	0.5.2-incubating	Biblioteca de pacotes para executar o Presto com o Hudi.
hue-server	4.6.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	1.1.0	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.7.0-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mahout-client	0.13.0	A biblioteca de Machine Learning.

Componente	Version (Versão)	Descrição
mxnet	1.5.1	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mariadb-server	5.5.64+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.2.88	Drivers NVIDIA e toolkit CUDA
oozie-client	5.2.0	O cliente da linha de comando do Oozie.
oozie-server	5.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	3.4.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.3- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.3- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.232	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.232	O serviço que executa partes de uma consulta.

Componente	Version (Versão)	Descrição
presto-client	0.232	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.3	O projeto R para computação estatística
ranger-kms-server	1.2.0	Sistema de gerenciamento de chaves Apache Ranger
spark-client	2.4.5-amzn-0.1	Os clientes da linha de comando do Spark.
spark-history-server	2.4.5-amzn-0.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.4.5-amzn-0.1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.4.5-amzn-0.1	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.

Componente	Version (Versão)	Descrição
tensorflow	1.14.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.2	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.8.2	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.14	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.14	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.30.2

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.30.2

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.

Classificações	Descrição
container-log4j	Altere os valores no arquivo container-log4j.properties YARN do Hadoop.
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4j-yarn-session.properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.

Classificações	Descrição
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.

Classificações	Descrição
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
https-env	Mude os valores no HTTPS ambiente.
https-site	Alterar os valores no arquivo https-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente do HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
hudi-env	Altere os valores no ambiente do Hudi.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.

Classificações	Descrição
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo <code>jupyterhub_config.py</code> do.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.
jupyter-sparkmagic-conf	Altere os valores no arquivo <code>config.json</code> do Sparkmagic.
livy-conf	Alterar os valores no arquivo <code>livy.conf</code> do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de <code>log4j.properties</code> no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo <code>mapred-site.xml</code> do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo <code>oozie-log4j.properties</code> do Oozie.
oozie-site	Alterar os valores no arquivo <code>oozie-site.xml</code> do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo <code>hadoop-metrics2-hbase.properties</code> do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo <code>hbase-site.xml</code> do Phoenix.
phoenix-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Phoenix.

Classificações	Descrição
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.

Classificações	Descrição
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.

Classificações	Descrição
ranger-kms-env	Altere os valores no KMS ambiente Ranger.
ranger-kms-log4j	Altere os valores no arquivo kms-log4j.properties do Ranger. KMS
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.
recordserver-env	Mude os valores no EMR RecordServer ambiente.
recordserver-conf	Altere os valores no arquivo EMR RecordServer erver.properties do.
recordserver-log4j	Altere os valores no EMR RecordServer arquivo log4j.properties.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.

Classificações	Descrição
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.30.1

Versões 5.30.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.30.1	emr-5.30.0	emr-5.29.0	emr-5.28.1
AWS SDK para Java	1.11.759	1.11.759	1.11.682	1.11.659
Python	2,7, 3,7	2,7, 3,7	2,7, 3,6	2,7, 3,6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
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
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

	emr-5.30.1	emr-5.30.0	emr-5.29.0	emr-5.28.1
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
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

Notas da versão 5.30.1

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.30.1 da Amazon. As alterações são referentes à versão 5.30.0.

Data da versão inicial: 30 de junho de 2020

Data da última atualização: 24 de agosto de 2020

Alterações, melhorias e problemas resolvidos

- As EMR versões mais recentes da Amazon corrigem o problema com um limite menor de “Máximo de arquivos abertos” em relação aos mais antigos AL2 na AmazonEMR. As EMR versões 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 e posteriores agora incluem uma correção permanente com uma configuração mais alta de “Máximo de arquivos abertos”.

- Corrigido um problema em que o processo do controlador da instância gerava um número infinito de processos.
- Corrigido um problema em que o Hue não conseguia executar uma consulta do Hive, mostrando a mensagem “o banco de dados está bloqueado” e impedindo a execução de consultas.
- Foi corrigido um problema do Spark para permitir que mais tarefas fossem executadas simultaneamente no EMR cluster.
- Corrigido um problema no caderno Jupyter que causava um “erro de muitos arquivos abertos” no servidor Jupyter.
- Corrigido um problema com as horas de início do cluster.

Novos atributos

- As interfaces de aplicativos persistentes do Tez UI e do YARN Timeline Server estão disponíveis com EMR as versões 6.x e 5.30.1 e EMR posteriores da Amazon. O acesso por link com um clique ao histórico persistente do aplicativo permite que você acesse rapidamente o histórico de tarefas sem configurar um proxy da web por meio de uma SSH conexão. Os logs de clusters ativos e encerrados ficam disponíveis por 30 dias após o término da aplicação. Para obter mais informações, consulte [Exibir interfaces de usuário de aplicativos persistentes](#) no Amazon EMR Management Guide.
- EMR execução do notebook APIs está disponível para executar EMR notebooks por meio de um script ou linha de comando. A capacidade de iniciar, parar, listar e descrever as execuções do EMR notebook sem o AWS console permite que você controle programaticamente um notebook. EMR Ao usar uma célula do caderno parametrizada, você pode passar valores de parâmetros diferentes para um caderno sem precisar criar uma cópia do caderno para cada novo conjunto de valores de parâmetros. Consulte [EMR APIs](#). Para obter um exemplo de código, consulte [Exemplos de comandos para executar EMR Notebooks programaticamente](#).

Problemas conhecidos

- Limite inferior de “Máximo de arquivos abertos” em versões mais antigas AL2 [corrigido em versões mais recentes]. EMR Lançamentos da Amazon: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 e emr-6.2.0 são baseados em versões mais antigas do Linux of Amazon 2 (AL2, que têm uma configuração de limite inferior para “Máximo de arquivos abertos” quando

os clusters da Amazon são criados com o padrão. EMR AMI As EMR versões 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 e posteriores incluem uma correção permanente com uma configuração maior de “Máximo de arquivos abertos”. Versões com o limite inferior de arquivos abertos causam o erro “Muitos arquivos abertos” ao ser enviado um trabalho do Spark. Nas versões afetadas, o padrão da Amazon AMI tem uma configuração EMR padrão de ulimit de 4096 para “Máximo de arquivos abertos”, que é inferior ao limite de 65536 arquivos no Linux 2. latestAmazon AMI A configuração inferior de ulimit para “Máximo de arquivos abertos” causa falhas em trabalhos do Spark quando o driver e o executor do Spark tentam abrir mais de 4096 arquivos. Para corrigir o problema, a Amazon EMR tem um script de ação de bootstrap (BA) que ajusta a configuração ulimit na criação do cluster.

Se você estiver usando uma EMR versão mais antiga da Amazon que não tem a solução permanente para esse problema, a solução alternativa a seguir permite que você defina explicitamente o ulimit do controlador de instância para um máximo de 65536 arquivos.

Defina explicitamente um ulimit na linha de comando

1. Edite `/etc/systemd/system/instance-controller.service` para adicionar os seguintes parâmetros à seção Serviço.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Reiniciar InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Defina um ulimit usando a ação de bootstrap (BA)

Você também pode usar um script de ação de bootstrap (BA) para configurar o ulimit instance-controller para 65536 arquivos na criação do 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
```

- EMRCadernos

O recurso que permite instalar kernels e bibliotecas Python adicionais no nó primário do cluster está desativado por padrão na versão 5.30.1. EMR Para obter mais informações sobre esse atributo, consulte [Instalar kernels e bibliotecas Python em um nó primário do cluster](#).

Para habilitar o recurso, faça o seguinte:

1. Certifique-se de que a política de permissões anexada à função de serviço para EMR Notebooks permita a seguinte ação:

```
elasticmapreduce:ListSteps
```

Para obter mais informações, consulte [Função de serviço para EMR notebooks](#).

2. Use o AWS CLI para executar uma etapa no cluster que configura os EMR Notebooks, conforme mostrado no exemplo a seguir. Substituir *us-east-1* com a região em que seu cluster reside. Para obter mais informações, consulte [Adding Steps to a Cluster Using the 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"]
```

- Ajuste de escala gerenciado

As operações de ajuste de escala gerenciado nos clusters das versões 5.30.0 e 5.30.1 sem o Presto instalado podem causar falhas na aplicação ou fazer com que um grupo de instâncias

ou uma frota de instâncias uniforme permaneça no estado `ARRESTED`, sobretudo quando uma operação de redução da escala verticalmente logo é seguida por uma operação de aumento da escala verticalmente.

Como solução alternativa, escolha o Presto como um aplicativo a ser instalado ao criar um cluster com as EMR versões 5.30.0 e 5.30.1 da Amazon, mesmo que seu trabalho não exija o Presto.

- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSH como usuário `hadoop` do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do `hadoop`.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo `keytab` está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

- Quando você usa o Spark com a formatação de localização de partições do Hive para ler dados no Amazon S3 e executa o Spark nas EMR versões 5.30.0 a 5.36.0 e 6.2.0 a 6.9.0 da Amazon, você pode encontrar um problema que impede seu cluster de ler os dados corretamente. Isso poderá acontecer se suas partições tiverem todas as características a seguir:
 - Duas ou mais partições são verificadas na mesma tabela.

- Pelo menos um caminho de diretório de partição é um prefixo de pelo menos outro caminho de diretório de partição, por exemplo, `s3://bucket/table/p=a` é um prefixo de `s3://bucket/table/p=a b`.
- O primeiro caractere que segue o prefixo no outro diretório de partição tem um valor UTF -8 menor que o `/` caractere (U+002F). Por exemplo, o caractere de espaço (U+0020) que ocorre entre `a` e `b` em `s3://bucket/table/p=a b` se enquadra nessa categoria. Observe que existem 14 outros caracteres que não são de controle: `!"#$%&'()*+,-.` Para obter mais informações, consulte a [tabela de codificação UTF -8 e caracteres Unicode](#).

Como solução alternativa para esse problema, defina a configuração `spark.sql.sources.fastS3PartitionDiscovery.enabled` como `false` na classificação `spark-defaults`.

Versões de componente 5.30.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.3.0	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.14.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
emr-goodies	2.13.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.5.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.14.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	1.5.0	EMRConector S3Select
emrfs	2.40.0	O conector do Amazon S3 para aplicativos do ecossistema do Hadoop.
flink-client	1.10.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.

Componente	Version (Versão)	Descrição
hadoop-client	2.8.5-amzn-6	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-6	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.8.5-amzn-6	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.8.5-amzn-6	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	2.8.5-amzn-6	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httpfs-server	2.8.5-amzn-6	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.8.5-amzn-6	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-6	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.8.5-amzn-6	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.8.5-amzn-6	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.

Componente	Version (Versão)	Descrição
hadoop-yarn-timeline-server	2.8.5-amzn-6	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.13	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.13	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.4.13	HBase cliente de linha de comando.
hbase-rest-server	1.4.13	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.4.13	Serviço que fornece um endpoint Thrift para HBase.
hcatalog-client	2.3.6-amzn-2	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.6-amzn-2	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.6-amzn-2	HTTP endpoint fornecendo uma REST interface para HCatalog.

Componente	Version (Versão)	Descrição
hive-client	2.3.6-amzn-2	O cliente da linha de comando do Hive.
hive-hbase	2.3.6-amzn-2	Cliente do Hive-hbase.
hive-metastore-server	2.3.6-amzn-2	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.6-amzn-2	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.5.2-incubating	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.
hudi-presto	0.5.2-incubating	Biblioteca de pacotes para executar o Presto com o Hudi.
hue-server	4.6.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	1.1.0	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.7.0-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso

Componente	Version (Versão)	Descrição
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.5.1	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mariadb-server	5.5.64	Meu servidor SQL de banco de dados.
nvidia-cuda	9.2.88	Drivers NVIDIA e toolkit CUDA
oozie-client	5.2.0	O cliente da linha de comando do Oozie.
oozie-server	5.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	3.4.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.3- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.3- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.232	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.232	O serviço que executa partes de uma consulta.

Componente	Version (Versão)	Descrição
presto-client	0.232	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.3	O projeto R para computação estatística
ranger-kms-server	1.2.0	Sistema de gerenciamento de chaves Apache Ranger
spark-client	2.4.5-amzn-0	Os clientes da linha de comando do Spark.
spark-history-server	2.4.5-amzn-0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.4.5-amzn-0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.4.5-amzn-0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.

Componente	Version (Versão)	Descrição
tensorflow	1.14.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.2	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.8.2	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.14	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.14	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.30.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.30.1

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.

Classificações	Descrição
container-log4j	Altere os valores no arquivo container-log4j.properties YARN do Hadoop.
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4j-yarn-session.properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.

Classificações	Descrição
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.

Classificações	Descrição
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
https-env	Mude os valores no HTTPFS ambiente.
https-site	Alterar os valores no arquivo https-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente do HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
hudi-env	Altere os valores no ambiente do Hudi.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.

Classificações	Descrição
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo <code>jupyterhub_config.py</code> do.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.
jupyter-sparkmagic-conf	Altere os valores no arquivo <code>config.json</code> do Sparkmagic.
livy-conf	Alterar os valores no arquivo <code>livy.conf</code> do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de <code>log4j.properties</code> no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo <code>mapred-site.xml</code> do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo <code>oozie-log4j.properties</code> do Oozie.
oozie-site	Alterar os valores no arquivo <code>oozie-site.xml</code> do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo <code>hadoop-metrics2-hbase.properties</code> do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo <code>hbase-site.xml</code> do Phoenix.
phoenix-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Phoenix.

Classificações	Descrição
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.

Classificações	Descrição
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.

Classificações	Descrição
ranger-kms-env	Altere os valores no KMS ambiente Ranger.
ranger-kms-log4j	Altere os valores no arquivo kms-log4j.properties do Ranger. KMS
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.
recordserver-env	Mude os valores no EMR RecordServer ambiente.
recordserver-conf	Altere os valores no arquivo EMR RecordServer erver.properties do.
recordserver-log4j	Altere os valores no EMR RecordServer arquivo log4j.properties.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.

Classificações	Descrição
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.30.0

Versões 5.30.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.30.0	emr-5.29.0	emr-5.28.1	emr-5.28.0
AWS SDK para Java	1.11.759	1.11.682	1.11.659	1.11.659
Python	2,7, 3,7	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
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
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

	emr-5.30.0	emr-5.29.0	emr-5.28.1	emr-5.28.0
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

Notas da versão 5.30.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.30.0 da Amazon. As alterações são referentes à versão 5.29.0.

Data da versão inicial: 13 de maio de 2020

Data da última atualização: 25 de junho de 2020

Atualizações

- Atualizado para AWS SDK for Java a versão 1.11.759
- Amazon SageMaker Spark atualizado SDK para a versão 1.3.0
- Servidor de EMR registros atualizado para a versão 1.6.0

- Atualização do Flink para a versão 1.10.0
- Atualização do Ganglia para a versão 3.7.2
- Atualizado HBase para a versão 1.4.13
- Atualização do Hudi para a versão 0.5.2-incubating
- Atualização do Hue para a versão 4.6.0
- Atualizado JupyterHub para a versão 1.1.0
- Atualização do Livy para a versão 0.7.0-incubating
- Atualização do Oozie para a versão 5.2.0
- Atualização do Presto para a versão 0.232
- Atualização do Spark para a versão 2.4.5
- Conectores e drivers atualizados: Amazon Glue Connector 1.12.0; Amazon Kinesis Connector 3.5.0; DynamoDB Connector 4.14.0 EMR

Novos atributos

- EMRNotebooks — Quando usados com EMR clusters criados usando 5.30.0, os kernels do EMR notebook são executados no cluster. Isso melhora o desempenho do bloco de anotações e permite que instalar e personalizar kernels. Você também pode instalar bibliotecas Python no nó primário do cluster. Para obter mais informações, consulte [Instalando e usando kernels e bibliotecas](#) no Guia EMRde gerenciamento.
- Escalabilidade gerenciada — Com a Amazon EMR versão 5.30.0 e posterior, você pode habilitar a escalabilidade EMR gerenciada para aumentar ou diminuir automaticamente o número de instâncias ou unidades em seu cluster com base na carga de trabalho. A Amazon avalia EMR continuamente as métricas do cluster para tomar decisões de escalabilidade que otimizem seus clusters em termos de custo e velocidade. Para obter mais informações, consulte [Scaling Cluster Resources](#) no Amazon EMR Management Guide.
- Criptografe arquivos de log armazenados no Amazon S3 — Com a EMR Amazon versão 5.30.0 e posterior, você pode criptografar arquivos de log armazenados no Amazon S3 com uma chave gerenciada pelo cliente. AWS KMS Para obter mais informações, consulte [Criptografar arquivos de log armazenados no Amazon S3](#) no EMRAmazon Management Guide.
- Suporte ao Amazon Linux 2 — Na EMR versão 5.30.0 e posterior, EMR usesAmazon sistema operacional Linux 2. A nova personalização AMIs (Amazon Machine Image) deve ser baseada no theAmazon Linux 2AMI. Para obter mais informações, consulte [Usando um personalizado AMI](#).

- Presto Graceful Auto Scale — EMR clusters que usam 5.30.0 podem ser configurados com um período de tempo limite de escalonamento automático que permite que as tarefas do Presto terminem de ser executadas antes que seu nó seja desativado. Para ter mais informações, consulte [Usar a escalabilidade automática do Presto com desativação tranquila](#).
- Criação de instância de frota com nova opção de estratégia de alocação — Uma nova opção de estratégia de alocação está disponível na EMR versão 5.12.1 e posterior. Ele oferece provisionamento de cluster mais rápido, alocação de spot mais precisa e menos interrupção de instâncias spot. Atualizações para funções de EMR serviço não padrão são necessárias. Consulte [Configurar frotas de instâncias](#).
- comandos `sudo systemctl stop` e `sudo systemctl start` — Na EMR versão 5.30.0 e posterior, qual sistema operacional use Amazon Linux 2 usa e comanda para reiniciar serviços. `EMR sudo systemctl stop sudo systemctl start` Para obter mais informações, consulte [Como reinício um serviço na AmazonEMR?](#).

Alterações, melhorias e problemas resolvidos

- EMR versão 5.30.0 não instala o Ganglia por padrão. É possível selecionar explicitamente o Ganglia para ser instalado ao criar um cluster.
- Otimizações do desempenho do Spark
- Otimizações do desempenho do Presto
- O Python 3 é o padrão para a Amazon EMR versão 5.30.0 e posterior.
- O grupo de segurança gerenciado padrão para acesso ao serviço em sub-redes privadas foi atualizado com novas regras. Se você usar um grupo de segurança personalizado para acesso ao serviço, será necessário incluir as mesmas regras do grupo de segurança gerenciado padrão. Para obter mais informações, consulte [Grupo EMR de segurança gerenciado pela Amazon para acesso a serviços \(sub-redes privadas\)](#). Se você usa uma função de serviço personalizada para a AmazonEMR, deve conceder permissão para que ela `ec2:describeSecurityGroups` EMR possa validar se os grupos de segurança foram criados corretamente. Se você usar o `EMR_DefaultRole`, essa permissão já estará incluída na política gerenciada padrão.

Problemas conhecidos

- Limite inferior de “Máximo de arquivos abertos” em versões mais antigas AL2 [corrigido em versões mais recentes]. EMRLançamentos da Amazon: `emr-5.30.x`, `emr-5.31.0`, `emr-5.32.0`, `emr-6.0.0`, `emr-6.1.0` e `emr-6.2.0` são baseados em versões mais antigas do Linux of Amazon 2

()AL2, que têm uma configuração de limite inferior para “Máximo de arquivos abertos” quando os clusters da Amazon são criados com o padrão. EMR AMI As EMR versões 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 e posteriores incluem uma correção permanente com uma configuração maior de “Máximo de arquivos abertos”. Versões com o limite inferior de arquivos abertos causam o erro “Muitos arquivos abertos” ao ser enviado um trabalho do Spark. Nas versões afetadas, o padrão da Amazon AMI tem uma configuração EMR padrão de `ulimit` de 4096 para “Máximo de arquivos abertos”, que é inferior ao limite de 65536 arquivos no Linux 2. latestAmazon AMI A configuração inferior de `ulimit` para “Máximo de arquivos abertos” causa falhas em trabalhos do Spark quando o driver e o executor do Spark tentam abrir mais de 4096 arquivos. Para corrigir o problema, a Amazon EMR tem um script de ação de bootstrap (BA) que ajusta a configuração `ulimit` na criação do cluster.

Se você estiver usando uma EMR versão mais antiga da Amazon que não tem a solução permanente para esse problema, a solução alternativa a seguir permite que você defina explicitamente o `ulimit` do controlador de instância para um máximo de 65536 arquivos.

Defina explicitamente um `ulimit` na linha de comando

1. Edite `/etc/systemd/system/instance-controller.service` para adicionar os seguintes parâmetros à seção Serviço.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Reiniciar InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Defina um `ulimit` usando a ação de bootstrap (BA)

Você também pode usar um script de ação de bootstrap (BA) para configurar o `ulimit instance-controller` para 65536 arquivos na criação do 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
```

- Ajuste de escala gerenciado

As operações de ajuste de escala gerenciado nos clusters das versões 5.30.0 e 5.30.1 sem o Presto instalado podem causar falhas na aplicação ou fazer com que um grupo de instâncias ou uma frota de instâncias uniforme permaneça no estado ARRESTED, sobretudo quando uma operação de redução da escala verticalmente logo é seguida por uma operação de aumento da escala verticalmente.

Como solução alternativa, escolha o Presto como um aplicativo a ser instalado ao criar um cluster com as EMR versões 5.30.0 e 5.30.1 da Amazon, mesmo que seu trabalho não exija o Presto.

- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSH como hadoop usuário do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do hadoop.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo `keytab` está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

- O mecanismo de banco de dados padrão para o Hue 4.6.0 é SQLite, o que causa problemas quando você tenta usar o Hue com um banco de dados externo. Para corrigir isso, defina `engine` na sua classificação de configuração `hue.ini` como `mysql`. Esse problema foi corrigido na EMR versão 5.30.1 da Amazon.
- Quando você usa o Spark com a formatação de localização de partições do Hive para ler dados no Amazon S3 e executa o Spark nas EMR versões 5.30.0 a 5.36.0 e 6.2.0 a 6.9.0 da Amazon, você pode encontrar um problema que impede seu cluster de ler os dados corretamente. Isso poderá acontecer se suas partições tiverem todas as características a seguir:
 - Duas ou mais partições são verificadas na mesma tabela.
 - Pelo menos um caminho de diretório de partição é um prefixo de pelo menos outro caminho de diretório de partição, por exemplo, `s3://bucket/table/p=a` é um prefixo de `s3://bucket/table/p=a b`.
 - O primeiro caractere que segue o prefixo no outro diretório de partição tem um valor UTF-8 menor que o `/` caractere (U+002F). Por exemplo, o caractere de espaço (U+0020) que ocorre entre `a` e `b` em `s3://bucket/table/p=a b` se enquadra nessa categoria. Observe que existem 14 outros caracteres que não são de controle: `!"#$%&'()*+,-.` Para obter mais informações, consulte a [tabela de codificação UTF-8 e caracteres Unicode](#).

Como solução alternativa para esse problema, defina a configuração `spark.sql.sources.fastS3PartitionDiscovery.enabled` como `false` na classificação `spark-defaults`.

Versões de componente 5.30.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR

e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.3.0	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.14.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.13.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.5.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-notebook-env</code>	1.0.0	Conda env para bloco de anotações do EMR
<code>emr-s3-dist-cp</code>	2.14.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
<code>emr-s3-select</code>	1.5.0	EMRConector S3Select
<code>emrfs</code>	2.40.0	O conector do Amazon S3 para aplicativos do ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
flink-client	1.10.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.5-amzn-6	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-6	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.8.5-amzn-6	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.8.5-amzn-6	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	2.8.5-amzn-6	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.

Componente	Version (Versão)	Descrição
hadoop-https-server	2.8.5-amzn-6	HTTP endpoint para HDFS operações.
hadoop-kms-server	2.8.5-amzn-6	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-6	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.8.5-amzn-6	YARN serviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.8.5-amzn-6	YARN serviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.5-amzn-6	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.13	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.13	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.4.13	HBase cliente de linha de comando.

Componente	Version (Versão)	Descrição
hbase-rest-server	1.4.13	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.4.13	Serviço que fornece um endpoint Thrift para HBase
hcatalog-client	2.3.6-amzn-2	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.6-amzn-2	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.6-amzn-2	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.3.6-amzn-2	O cliente da linha de comando do Hive.
hive-hbase	2.3.6-amzn-2	Cliente do Hive-hbase.
hive-metastore-server	2.3.6-amzn-2	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.6-amzn-2	O serviço que aceita as consultas do Hive como solicitações da web.

Componente	Version (Versão)	Descrição
hudi	0.5.2-incubating	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.
hudi-presto	0.5.2-incubating	Biblioteca de pacotes para executar o Presto com o Hudi.
hue-server	4.6.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	1.1.0	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.7.0-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.5.1	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mariadb-server	5.5.64	Meu servidor SQL de banco de dados.
nvidia-cuda	9.2.88	Drivers NVIDIA e toolkit CUDA
oozie-client	5.2.0	O cliente da linha de comando do Oozie.

Componente	Version (Versão)	Descrição
oozie-server	5.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	3.4.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.3- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.3- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.232	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.232	O serviço que executa partes de uma consulta.
presto-client	0.232	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.3	O projeto R para computação estatística

Componente	Version (Versão)	Descrição
ranger-kms-server	1.2.0	Sistema de gerenciamento de chaves Apache Ranger
spark-client	2.4.5-amzn-0	Os clientes da linha de comando do Spark.
spark-history-server	2.4.5-amzn-0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.4.5-amzn-0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.4.5-amzn-0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	1.14.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.2	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.8.2	O notebook baseado na web que permite a análise de dados interativa.

Componente	Version (Versão)	Descrição
zookeeper-server	3.4.14	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.14	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.30.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.30.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.

Classificações	Descrição
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.

Classificações	Descrição
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.

Classificações	Descrição
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente do HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
hudi-env	Altere os valores no ambiente do Hudi.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.

Classificações	Descrição
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.

Classificações	Descrição
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.

Classificações	Descrição
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.
ranger-kms-log4j	Altere os valores no arquivo kms-log4j.properties do Ranger. KMS
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.
recordserver-env	Mude os valores no EMR RecordServer ambiente.
recordserver-conf	Altere os valores no arquivo EMR RecordServererver.properties do.

Classificações	Descrição
recordserver-log4j	Altere os valores no EMR RecordServer arquivo log4j.properties.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.

Classificações	Descrição
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.29.0

Versões 5.29.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.29.0	emr-5.28.1	emr-5.28.0	emr-5.27.1
AWS SDK para Java	1.11.682	1.11.659	1.11.659	1.11.615
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2.11.12	2.11.12	2.11.12	2.11.12

	emr-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

	emr-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

Notas da versão 5.29.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.29.0 da Amazon. As alterações são referentes à versão 5.28.1.

Data da versão inicial: 17 de janeiro de 2020

Atualizações

- Atualizado para AWS SDK for Java a versão 1.11.682
- Atualizado o Hive para a versão 2.3.6
- Atualizado o Flink para a versão 1.9.1
- Atualizado o EMRFS para a versão 2.38.0
- Conector EMR DynamoDB atualizado para a versão 4.13.0

Alterações, melhorias e problemas resolvidos

- Spark
 - Otimizações do desempenho do Spark

- EMRFS

- O Guia de gerenciamento é atualizado para as configurações padrão `emrfs-site.xml` para uma visualização consistente.

Problemas conhecidos

- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSH como `hadoop` usuário do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do `hadoop`.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo `keytab` está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

Versões de componente 5.29.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR

e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.2.6	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.13.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.12.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.13.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
<code>emr-s3-select</code>	1.4.0	EMRConector S3Select
<code>emrfs</code>	2.38.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
<code>flink-client</code>	1.9.1	Scripts do cliente da linha de comando e aplicativos do Apache Flink.

Componente	Version (Versão)	Descrição
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.5-amzn-5	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-5	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.8.5-amzn-5	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.8.5-amzn-5	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	2.8.5-amzn-5	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httpfs-server	2.8.5-amzn-5	HTTPendpoint para HDFS operações.

Componente	Version (Versão)	Descrição
hadoop-kms-server	2.8.5-amzn-5	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-5	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.8.5-amzn-5	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.8.5-amzn-5	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.5-amzn-5	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.10	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.10	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.4.10	HBasecliente de linha de comando.
hbase-rest-server	1.4.10	Serviço que fornece um RESTful HTTP endpoint paraHBase.

Componente	Version (Versão)	Descrição
hbase-thrift-server	1.4.10	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	2.3.6-amzn-1	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.6-amzn-1	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.6-amzn-1	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	2.3.6-amzn-1	O cliente da linha de comando do Hive.
hive-hbase	2.3.6-amzn-1	Cliente do Hive-hbase.
hive-metastore-server	2.3.6-amzn-1	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.6-amzn-1	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.5.0-incubating	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.

Componente	Version (Versão)	Descrição
hudi-presto	0.5.0-incubating	Biblioteca de pacotes para executar o Presto com o Hudi.
hue-server	4.4.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	1.0.0	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.6.0-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.5.1	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.2.88	Drivers NVIDIA e toolkit CUDA
oozie-client	5.1.0	O cliente da linha de comando do Oozie.
oozie-server	5.1.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.

Componente	Version (Versão)	Descrição
opencv	3.4.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.3- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.3- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.227	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.227	O serviço que executa partes de uma consulta.
presto-client	0.227	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.1	O projeto R para computação estatística
spark-client	2.4.4	Os clientes da linha de comando do Spark.

Componente	Version (Versão)	Descrição
spark-history-server	2.4.4	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.4.4	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.4.4	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	1.14.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.2	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.
zeppelin-server	0.8.2	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.14	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.

Componente	Version (Versão)	Descrição
zookeeper-client	3.4.14	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.29.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.29.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Alterar as configurações de <code>log4j-cli.properties</code> no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.

Classificações	Descrição
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.

Classificações	Descrição
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.

Classificações	Descrição
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente do HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.

Classificações	Descrição
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.

Classificações	Descrição
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.

Classificações	Descrição
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.
ranger-kms-log4j	Altere os valores no arquivo kms-log4j.properties do Ranger. KMS
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.
recordserver-env	Mude os valores no EMR RecordServer ambiente.
recordserver-conf	Altere os valores no arquivo EMR RecordServer erver.properties do.
recordserver-log4j	Altere os valores no EMR RecordServer arquivo log4j.properties.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.

Classificações	Descrição
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.28.1

Versões 5.28.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.28.1	emr-5.28.0	emr-5.27.1	emr-5.27.0
AWS SDK para Java	1.11.659	1.11.659	1.11.615	1.11.615
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
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

	emr-5.28.1	emr-5.28.0	emr-5.27.1	emr-5.27.0
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
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

	emr-5.28.1	emr-5.28.0	emr-5.27.1	emr-5.27.0
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

Notas da versão 5.28.1

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.28.1 da Amazon. As alterações são referentes à versão 5.28.0.

Data da versão inicial: 10 de janeiro de 2020

Alterações, melhorias e problemas resolvidos

- Spark
 - Correção de problemas de compatibilidade do Spark.
- CloudWatch Métricas
 - Foi corrigida a publicação do Amazon CloudWatch Metrics em um EMR cluster com vários nós primários.
- Desabilitada mensagem de log
 - Desabilitada mensagem de log falsa, "... uso de versão antiga (<4.5.8) do cliente Apache http".

Problemas conhecidos

- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSH como usuário do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do hadoop.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo keytab está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

Versões de componente 5.28.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
aws-sagemaker-spark-sdk	1.2.6	Amazon SageMaker Spark SDK

Componente	Version (Versão)	Descrição
emr-ddb	4.12.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.11.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.13.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	1.3.0	EMRConector S3Select
emrfs	2.37.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.9.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.

Componente	Version (Versão)	Descrição
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.5-amzn-5	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-5	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.8.5-amzn-5	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.8.5-amzn-5	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	2.8.5-amzn-5	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httpfs-server	2.8.5-amzn-5	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.8.5-amzn-5	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-5	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.

Componente	Version (Versão)	Descrição
hadoop-yarn-nodemanager	2.8.5-amzn-5	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.8.5-amzn-5	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.5-amzn-5	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.10	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.10	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.4.10	HBasecliente de linha de comando.
hbase-rest-server	1.4.10	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	1.4.10	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	2.3.6-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.

Componente	Version (Versão)	Descrição
hcatalog-server	2.3.6-amzn-0	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.6-amzn-0	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	2.3.6-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.3.6-amzn-0	Cliente do Hive-hbase.
hive-metastore-server	2.3.6-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.6-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.5.0-incubating	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.
hudi-presto	0.5.0-incubating	Biblioteca de pacotes para executar o Presto com o Hudi.
hue-server	4.4.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop

Componente	Version (Versão)	Descrição
jupyterhub	1.0.0	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.6.0-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.5.1	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.2.88	Drivers NVIDIA e toolkit CUDA
oozie-client	5.1.0	O cliente da linha de comando do Oozie.
oozie-server	5.1.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	3.4.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.3- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente

Componente	Version (Versão)	Descrição
phoenix-query-server	4.14.3- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.227	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.227	O serviço que executa partes de uma consulta.
presto-client	0.227	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.1	O projeto R para computação estatística
spark-client	2.4.4	Os clientes da linha de comando do Spark.
spark-history-server	2.4.4	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.4.4	Mecanismo de execução na memória para YARN.

Componente	Version (Versão)	Descrição
spark-yarn-slave	2.4.4	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	1.14.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.2	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.8.2	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.14	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.14	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.28.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.28.1

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.
container-log4j	Altere os valores no arquivo container-log4j.properties YARN do Hadoop.
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4j-yarn-session.properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase

Classificações	Descrição
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.

Classificações	Descrição
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente do HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.

Classificações	Descrição
jupyter-notebook-conf	Alterar os valores no arquivo <code>jupyter_notebook_config.py</code> do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo <code>jupyterhub_config.py</code> do.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.
jupyter-sparkmagic-conf	Altere os valores no arquivo <code>config.json</code> do Sparkmagic.
livy-conf	Alterar os valores no arquivo <code>livy.conf</code> do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de <code>log4j.properties</code> no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo <code>mapred-site.xml</code> do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo <code>oozie-log4j.properties</code> do Oozie.
oozie-site	Alterar os valores no arquivo <code>oozie-site.xml</code> do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo <code>hadoop-metrics2-hbase.properties</code> do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo <code>hbase-site.xml</code> do Phoenix.

Classificações	Descrição
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.

Classificações	Descrição
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS

Classificações	Descrição
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.
ranger-kms-log4j	Altere os valores no arquivo kms-log4j.properties do Ranger. KMS
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.
recordserver-env	Mude os valores no EMR RecordServer ambiente.
recordserver-conf	Altere os valores no arquivo EMR RecordServer erver.properties do.
recordserver-log4j	Altere os valores no EMR RecordServer arquivo log4j.properties.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.

Classificações	Descrição
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.28.0

Versões 5.28.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hudi](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)

- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.28.0	emr-5.27.1	emr-5.27.0	emr-5.26.0
AWS SDK para Java	1.11.659	1.11.615	1.11.615	1.11.595
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
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	-	-	-	-

	emr-5.28.0	emr-5.27.1	emr-5.27.0	emr-5.26.0
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
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

Notas da versão 5.28.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.28.0 da Amazon. As alterações são referentes à 5.27.0.

Data da versão inicial: 12 de novembro de 2019

Atualizações

- Flink atualizado para a versão 1.9.0
- Atualizado o Hive para a versão 2.3.6
- Atualizado MXNet para a versão 1.5.1
- Phoenix atualizado para a versão 4.14.3
- Presto atualizado para a versão 0.227
- Zeppelin atualizado para a versão 0.8.2

Novos atributos

- [O Apache Hudi](#) agora está disponível EMR para a Amazon instalar quando você cria um cluster. Para ter mais informações, consulte [Hudi](#).
- (25 de novembro de 2019) Agora você pode optar por executar várias etapas em paralelo para melhorar a utilização do cluster e economizar custos. Pode também cancelar etapas pendentes e em execução. Para obter mais informações, consulte [Trabalhar com etapas usando o console AWS CLI e](#).
- (3 de dezembro de 2019) Agora você pode criar e executar EMR clusters no AWS Outposts. AWS Outposts habilita AWS serviços, infraestrutura e modelos operacionais nativos em instalações locais. Em AWS Outposts ambientes, você pode usar as mesmas AWS APIs ferramentas e infraestrutura que usa na AWS nuvem. Para obter mais informações, consulte [EMRclusters on AWS Outposts](#).
- (11 de março de 2020) A partir da EMR versão 5.28.0 da Amazon, você pode criar e executar EMR clusters da Amazon em uma sub-rede de Zonas AWS Locais como uma extensão lógica de uma AWS região que suporta Zonas Locais. Uma zona local permite que os EMR recursos da Amazon e um subconjunto de AWS serviços, como serviços de computação e armazenamento, estejam localizados mais perto dos usuários, fornecendo acesso de latência muito baixa a aplicativos executados localmente. Para obter uma lista das zonas locais disponíveis, consulte [Zonas locais da AWS](#). Para obter informações sobre como acessar as Zonas AWS Locais disponíveis, consulte [Regiões, Zonas de Disponibilidade e Zonas Locais](#).

No momento, as Zonas Locais não oferecem suporte a Amazon EMR Notebooks e não oferecem suporte a conexões diretamente com a Amazon EMR usando a interface VPC endpoint (AWS PrivateLink).

Alterações, melhorias e problemas resolvidos

- Suporte expandido do aplicativo para clusters de alta disponibilidade
 - Para obter mais informações, consulte [Aplicativos compatíveis em um EMR cluster com vários nós primários](#) no Amazon EMR Management Guide.
- Spark
 - Otimizações da performance
- Hive
 - Otimizações da performance
- Presto
 - Otimizações da performance

Problemas conhecidos

- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSH como usuário do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do hadoop.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo keytab está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

Versões de componente 5.28.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.2.6	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.12.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.11.0	Bibliotecas convenientes para o ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
emr-kinesis	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.13.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	1.3.0	EMRConector S3Select
emrfs	2.37.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.9.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.5-amzn-5	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.

Componente	Version (Versão)	Descrição
hadoop-hdfs-datanode	2.8.5-amzn-5	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.8.5-amzn-5	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.8.5-amzn-5	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	2.8.5-amzn-5	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-https-server	2.8.5-amzn-5	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.8.5-amzn-5	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-5	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.8.5-amzn-5	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.8.5-amzn-5	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.5-amzn-5	Serviço para recuperar informações atuais e históricas para YARN aplicativos.

Componente	Version (Versão)	Descrição
hbase-hmaster	1.4.10	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.10	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.4.10	HBase cliente de linha de comando.
hbase-rest-server	1.4.10	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.4.10	Serviço que fornece um endpoint Thrift para HBase.
hcatalog-client	2.3.6-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.6-amzn-0	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.6-amzn-0	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.3.6-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.3.6-amzn-0	Cliente do Hive-hbase.

Componente	Version (Versão)	Descrição
hive-metastore-server	2.3.6-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.6-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hudi	0.5.0-incubating	Estrutura de processamento incremental para alimentar o pipeline de dados com baixa latência e alta eficiência.
hudi-presto	0.5.0-incubating	Biblioteca de pacotes para executar o Presto com o Hudi.
hue-server	4.4.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	1.0.0	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.6.0-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mahout-client	0.13.0	A biblioteca de Machine Learning.

Componente	Version (Versão)	Descrição
mxnet	1.5.1	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.2.88	Drivers NVIDIA e toolkit CUDA
oozie-client	5.1.0	O cliente da linha de comando do Oozie.
oozie-server	5.1.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	3.4.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.3- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.3- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.227	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.227	O serviço que executa partes de uma consulta.

Componente	Version (Versão)	Descrição
presto-client	0.227	Cliente de linha de comando do Presto instalado em mestres de espera de um cluster de alta disponibilidade em que o servidor do Presto não é iniciado.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.1	O projeto R para computação estatística
spark-client	2.4.4	Os clientes da linha de comando do Spark.
spark-history-server	2.4.4	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.4.4	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.4.4	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	1.14.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.

Componente	Version (Versão)	Descrição
tez-on-yarn	0.9.2	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.8.2	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.14	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.14	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.28.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.28.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.

Classificações	Descrição
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.

Classificações	Descrição
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.

Classificações	Descrição
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
https-env	Mude os valores no HTTPFS ambiente.
https-site	Alterar os valores no arquivo https-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente do HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.

Classificações	Descrição
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.

Classificações	Descrição
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.

Classificações	Descrição
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.
ranger-kms-log4j	Altere os valores no arquivo kms-log4j.properties do Ranger. KMS
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.

Classificações	Descrição
recordserver-env	Mude os valores no EMR RecordServer ambiente.
recordserver-conf	Altere os valores no arquivo EMR RecordServer <code>erver.properties</code> do.
recordserver-log4j	Altere os valores no EMR RecordServer arquivo <code>log4j.properties</code> .
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo <code>spark-defaults.conf</code> do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo <code>hive-site.xml</code> do Spark.
spark-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Spark.
spark-metrics	Alterar os valores no arquivo <code>metrics.properties</code> do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo <code>oraoop-site.xml</code> OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo <code>sqoop-site.xml</code> do Sqoop.
tez-site	Alterar os valores no arquivo <code>tez-site.xml</code> do Tez.
yarn-env	Mude os valores no YARN ambiente.

Classificações	Descrição
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.27.1

Versões 5.27.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.27.1	emr-5.27.0	emr-5.26.0	emr-5.25.0
AWS SDK para Java	1.11.615	1.11.615	1.11.595	1.11.566

	emr-5.27.1	emr-5.27.0	emr-5.26.0	emr-5.25.0
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
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

	emr-5.27.1	emr-5.27.0	emr-5.26.0	emr-5.25.0
Phoenix	4.14.2	4.14.2	4.14.2	4.14.1
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

Notas da versão 5.27.1

Esta é uma versão de patch. Todos os aplicativos e componentes são iguais aos da EMR versão anterior da Amazon.

Status de suporte do Instance Metadata Service (IMDS) V2: os componentes Amazon EMR 5.23.1, 5.27.1 e 5.32 ou posteriores são usados para todas as chamadas. IMDSv2 IMDS Para IMDS chamadas no código do seu aplicativo, você pode usar ambos IMDSv1 e IMDSv2, ou configurar o IMDS para uso somente IMDSv2 para aumentar a segurança. Para outras EMR versões 5.x, a desativação IMDSv1 causa falha na inicialização do cluster.

Versões de componente 5.27.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são

a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion-amzn-EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado myapp-component com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como 2.2-amzn-2.

Componente	Version (Versão)	Descrição
aws-sagemaker-spark-sdk	1.2.4	Amazon SageMaker Spark SDK
emr-ddb	4.12.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.11.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.13.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	1.3.0	EMRConector S3Select
emrfs	2.36.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.8.1	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do

Componente	Version (Versão)	Descrição
		ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.5-amzn-4	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-4	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.8.5-amzn-4	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.8.5-amzn-4	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	2.8.5-amzn-4	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httpfs-server	2.8.5-amzn-4	HTTPendpoint para HDFS operações.

Componente	Version (Versão)	Descrição
hadoop-kms-server	2.8.5-amzn-4	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-4	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.8.5-amzn-4	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.8.5-amzn-4	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.5-amzn-4	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.10	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.10	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.4.10	HBasecliente de linha de comando.
hbase-rest-server	1.4.10	Serviço que fornece um RESTful HTTP endpoint paraHBase.

Componente	Version (Versão)	Descrição
hbase-thrift-server	1.4.10	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	2.3.5-amzn-1	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.5-amzn-1	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.5-amzn-1	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	2.3.5-amzn-1	O cliente da linha de comando do Hive.
hive-hbase	2.3.5-amzn-1	Cliente do Hive-hbase.
hive-metastore-server	2.3.5-amzn-1	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.5-amzn-1	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.4.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop

Componente	Version (Versão)	Descrição
jupyterhub	1.0.0	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.6.0-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.4.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.2.88	Drivers NVIDIA e toolkit CUDA
oozie-client	5.1.0	O cliente da linha de comando do Oozie.
oozie-server	5.1.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	3.4.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.2- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente

Componente	Version (Versão)	Descrição
phoenix-query-server	4.14.2- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.224	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.224	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.1	O projeto R para computação estatística
spark-client	2.4.4	Os clientes da linha de comando do Spark.
spark-history-server	2.4.4	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.4.4	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.4.4	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.

Componente	Version (Versão)	Descrição
tensorflow	1.14.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.2	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.8.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.14	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.14	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.27.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.27.1

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.

Classificações	Descrição
container-log4j	Altere os valores no arquivo container-log4j.properties YARN do Hadoop.
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4j-yarn-session.properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.

Classificações	Descrição
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.

Classificações	Descrição
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente do HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.

Classificações	Descrição
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.

Classificações	Descrição
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.

Classificações	Descrição
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.

Classificações	Descrição
ranger-kms-log4j	Altere os valores no arquivo kms-log4j.properties do Ranger. KMS
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.
recordserver-env	Mude os valores no EMR RecordServer ambiente.
recordserver-conf	Altere os valores no arquivo EMR RecordServer erver.properties do.
recordserver-log4j	Altere os valores no EMR RecordServer arquivo log4j.properties.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.

Classificações	Descrição
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.27.0

Versões 5.27.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livvy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.27.0	emr-5.26.0	emr-5.25.0	emr-5.24.1
AWS SDK para Java	1.11.615	1.11.595	1.11.566	1.11.546
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
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	-	-	-	-
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

	emr-5.27.0	emr-5.26.0	emr-5.25.0	emr-5.24.1
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	4.14.2	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

Notas da versão 5.27.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.27.0 da Amazon. As alterações são referentes à versão 5.26.0.

Data da versão inicial: 23 de setembro de 2019

Atualizações

- AWS SDK for Java 1.11.615
- Flink 1.8.1
- JupyterHub 1.0.0
- Spark 2.4.4

- Tensorflow 1.14.0
- Conectores e drivers:
 - Conector do DynamoDB 4.12.0

Novos atributos

- (24 de outubro de 2019) Os seguintes novos recursos em EMR notebooks estão disponíveis em todos os lançamentos da AmazonEMR.
 - Agora você pode associar repositórios Git a EMR notebooks para armazenar seus cadernos em um ambiente com controle de versão. Você pode compartilhar códigos com pares e reutilizar cadernos Jupyter existentes por meio de repositórios do Git remotos. Para obter mais informações, consulte [Associar repositórios Git aos Amazon EMR Notebooks no Guia de gerenciamento](#) da Amazon. EMR
 - O [utilitário nbdime](#) agora está disponível em EMR notebooks para simplificar a comparação e mesclagem de notebooks.
 - EMRnotebooks agora são compatíveis JupyterLab. JupyterLab é um ambiente de desenvolvimento interativo baseado na Web totalmente compatível com os notebooks Jupyter. Agora você pode optar por abrir seu caderno em qualquer um dos editores de cadernos JupyterLab ou no editor de cadernos Jupyter.
- (30 de outubro de 2019) Com EMR as versões 5.25.0 e posteriores da Amazon, você pode se conectar à interface do usuário do servidor de histórico do Spark na página de resumo do cluster ou na guia Histórico do aplicativo no console. Em vez de configurar um proxy web por meio de uma SSH conexão, você pode acessar rapidamente a interface do servidor de histórico do Spark para visualizar as métricas do aplicativo e acessar arquivos de log relevantes para clusters ativos e encerrados. Para obter mais informações, consulte [Acesso fora do cluster a interfaces de usuário de aplicativos persistentes](#) no Amazon EMR Management Guide.

Alterações, melhorias e problemas resolvidos

- EMRCluster Amazon com vários nós primários
 - Você pode instalar e executar o Flink em um EMR cluster da Amazon com vários nós primários. Para obter mais informações, consulte [Supported applications and features](#).
 - Você pode configurar a criptografia HDFS transparente em um EMR cluster da Amazon com vários nós primários. Para obter mais informações, consulte [Criptografia HDFS transparente em EMR clusters com vários nós primários](#).

- Agora você pode modificar a configuração dos aplicativos em execução em um EMR cluster da Amazon com vários nós primários. Para obter mais informações, consulte [Supplying a Configuration for an Instance Group in a Running Cluster](#).
- Conector Amazon EMR - DynamoDB
 - O Amazon EMR -DynamoDB Connector agora oferece suporte aos seguintes tipos de dados do DynamoDB: boolean, list, map, item, null. Para obter mais informações, consulte [Configurar uma tabela do Hive para executar comandos do Hive](#).

Problemas conhecidos

- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSH como hadoop usuário do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do hadoop.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo keytab está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

Versões de componente 5.27.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.2.4	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.12.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.11.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.13.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
<code>emr-s3-select</code>	1.3.0	EMRConector S3Select
<code>emrfs</code>	2.36.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
flink-client	1.8.1	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.5-amzn-4	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-4	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.8.5-amzn-4	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.8.5-amzn-4	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	2.8.5-amzn-4	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.

Componente	Version (Versão)	Descrição
hadoop-https-server	2.8.5-amzn-4	HTTP endpoint para HDFS operações.
hadoop-kms-server	2.8.5-amzn-4	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-4	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.8.5-amzn-4	YARN serviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.8.5-amzn-4	YARN serviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.5-amzn-4	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.10	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.10	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.4.10	HBase cliente de linha de comando.

Componente	Version (Versão)	Descrição
hbase-rest-server	1.4.10	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.4.10	Serviço que fornece um endpoint Thrift para HBase
hcatalog-client	2.3.5-amzn-1	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.5-amzn-1	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.5-amzn-1	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.3.5-amzn-1	O cliente da linha de comando do Hive.
hive-hbase	2.3.5-amzn-1	Cliente do Hive-hbase.
hive-metastore-server	2.3.5-amzn-1	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.5-amzn-1	O serviço que aceita as consultas do Hive como solicitações da web.

Componente	Version (Versão)	Descrição
hue-server	4.4.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	1.0.0	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.6.0-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.4.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.2.88	Drivers NVIDIA e toolkit CUDA
oozie-client	5.1.0	O cliente da linha de comando do Oozie.
oozie-server	5.1.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	3.4.0	Biblioteca de Visão Computacional de Código Aberto.

Componente	Version (Versão)	Descrição
phoenix-library	4.14.2- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.2- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.224	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.224	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.1	O projeto R para computação estatística
spark-client	2.4.4	Os clientes da linha de comando do Spark.
spark-history-server	2.4.4	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.4.4	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.4.4	Bibliotecas do Apache Spark necessárias para escravos. YARN

Componente	Version (Versão)	Descrição
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	1.14.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.2	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.8.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.14	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.14	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.27.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.27.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.
container-log4j	Altere os valores no arquivo container-log4j.properties YARN do Hadoop.
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4j-yarn-session.properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase

Classificações	Descrição
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.

Classificações	Descrição
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente do HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.

Classificações	Descrição
jupyter-notebook-conf	Alterar os valores no arquivo <code>jupyter_notebook_config.py</code> do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo <code>jupyterhub_config.py</code> do.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.
jupyter-sparkmagic-conf	Altere os valores no arquivo <code>config.json</code> do Sparkmagic.
livy-conf	Alterar os valores no arquivo <code>livy.conf</code> do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de <code>log4j.properties</code> no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo <code>mapred-site.xml</code> do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo <code>oozie-log4j.properties</code> do Oozie.
oozie-site	Alterar os valores no arquivo <code>oozie-site.xml</code> do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo <code>hadoop-metrics2-hbase.properties</code> do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo <code>hbase-site.xml</code> do Phoenix.

Classificações	Descrição
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.

Classificações	Descrição
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS

Classificações	Descrição
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.
ranger-kms-log4j	Altere os valores no arquivo kms-log4j.properties do Ranger. KMS
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.
recordserver-env	Mude os valores no EMR RecordServer ambiente.
recordserver-conf	Altere os valores no arquivo EMR RecordServer erver.properties do.
recordserver-log4j	Altere os valores no EMR RecordServer arquivo log4j.properties.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.

Classificações	Descrição
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.26.0

Versões 5.26.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)

- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.26.0	emr-5.25.0	emr-5.24.1	emr-5.24.0
AWS SDK para Java	1.11.595	1.11.566	1.11.546	1.11.546
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
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	-	-	-	-

	emr-5.26.0	emr-5.25.0	emr-5.24.1	emr-5.24.0
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
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

Notas da versão 5.26.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.26.0 da Amazon. As alterações são referentes à versão 5.25.0.

Data da versão inicial: 8 de agosto de 2019

Data da última atualização: 19 de agosto de 2019

Atualizações

- AWS SDK for Java 1.11.595
- HBase1.4.10
- Phoenix 4.14.2
- Conectores e drivers:
 - Conector do DynamoDB 4.11.0
 - Conector do MariaDB 2.4.2
 - Driver Amazon Redshift 1.2.32.1056 JDBC

Novos atributos

- (Beta) Com o Amazon EMR 5.26.0, você pode lançar um cluster que se integra ao Lake Formation. Essa integração fornece acesso refinado em nível de coluna a bancos de dados e tabelas no Glue Data Catalog. AWS Ele também permite o login único federado em EMR notebooks ou Apache Zeppelin a partir de um sistema de identidade corporativo. Para obter mais informações, consulte [Integração da Amazon EMR com AWS Lake Formation \(Beta\)](#).
- (19 de agosto de 2019) O acesso público de EMR bloqueio da Amazon agora está disponível com todos os EMR lançamentos da Amazon que oferecem suporte a grupos de segurança. Bloquear o acesso público é uma configuração de toda a conta aplicada a cada AWS região. Bloquear o acesso público impede que um cluster seja iniciado quando qualquer grupo de segurança associado ao cluster tem uma regra que permite tráfego de entrada de IPv4 0.0.0.0/0 ou IPv6:: /0 (acesso público) em uma porta, a menos que uma porta seja especificada como uma exceção. A porta 22 é uma exceção por padrão. Para obter mais informações, consulte [Usando o Amazon EMR Block Public Access](#) no Amazon EMR Management Guide.

Alterações, melhorias e problemas resolvidos

- Notebooks do EMR
 - Com a EMR versão 5.26.0 e versões posteriores, o EMR Notebooks oferece suporte a bibliotecas Python com escopo de notebook, além das bibliotecas Python padrão. Você pode instalar bibliotecas com escopo de caderno de dentro do editor de caderno sem precisar recriar um cluster ou reanexar um caderno a um cluster. As bibliotecas com escopo de caderno são criadas em um ambiente Python virtual para serem aplicadas somente à sessão de caderno

atual. Isso permite isolar as dependências do caderno. Para obter mais informações, consulte [Usando bibliotecas com escopo de notebook](#) no Amazon EMR Management Guide.

- EMRFS
 - Você pode ativar um recurso ETag de verificação (Beta) configurando `fs.s3.consistent.metadata.etag.validation.enabled` como `true`. Com esse recurso, EMRFS usa o Amazon S3 ETags para verificar se os objetos que estão sendo lidos são a versão mais recente disponível. Esse recurso é útil para casos de read-after-update uso em que os arquivos no Amazon S3 são sobrescritos, mantendo o mesmo nome. Atualmente, esse recurso de ETag verificação não funciona com o S3 Select. Para obter mais informações, consulte [Configurar visualização consistente](#).
- Spark
 - As seguintes otimizações agora estão habilitadas por padrão: remoção dinâmica de partições, DISTINCT antes, melhorias na inferência de estatísticas do SQL plano para consultas JOIN seguidas INTERSECT, nivelamento de DISTINCT subconsultas escalares, reordenamento otimizado de junções e junção com filtro bloom. Para obter mais informações, consulte [Otimizar a performance do Spark](#).
 - Aprimorada a geração de código de estágio completo para Sort-Merge Join.
 - Aprimorado o fragmento de consulta e a reutilização de subconsultas.
 - Melhorias na pré-alocação de executores na inicialização do Spark.
 - As junções com filtro de Bloom não são mais aplicadas quando o lado menor da junção inclui uma dica de transmissão.
- Tez
 - Resolvido um problema com o Tez. O Tez UI agora funciona em um EMR cluster da Amazon com vários nós primários.

Problemas conhecidos

- Os recursos aprimorados de geração de código em todo o estágio de Sort Merge Join podem aumentar a pressão de memória quando habilitados. Essa otimização melhora a performance, mas pode resultar em novas tentativas ou falhas de trabalho se `spark.yarn.executor.memoryOverheadFactor` não for ajustado para fornecer memória suficiente. Para desabilitar esse atributo, defina `spark.sql.sortMergeJoinExec.extendedCodegen.enabled` como `false`.
- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSH como hadoop usuário do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do hadoop.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo keytab está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

Versões de componente 5.26.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion-amzn-EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto

nomeado myapp-component com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como 2.2-amzn-2.

Componente	Version (Versão)	Descrição
aws-sagemaker-spark-sdk	1.2.4	Amazon SageMaker Spark SDK
emr-ddb	4.11.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.10.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.12.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	1.3.0	EMRConector S3Select
emrfs	2.35.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.8.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.

Componente	Version (Versão)	Descrição
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.5-amzn-4	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-4	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.8.5-amzn-4	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.8.5-amzn-4	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	2.8.5-amzn-4	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httpfs-server	2.8.5-amzn-4	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.8.5-amzn-4	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API

Componente	Version (Versão)	Descrição
hadoop-mapred	2.8.5-amzn-4	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.8.5-amzn-4	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.8.5-amzn-4	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.5-amzn-4	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.10	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.10	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.4.10	HBasecliente de linha de comando.
hbase-rest-server	1.4.10	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	1.4.10	Serviço que fornece um endpoint Thrift para. HBase

Componente	Version (Versão)	Descrição
hcatalog-client	2.3.5-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.5-amzn-0	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.5-amzn-0	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	2.3.5-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.3.5-amzn-0	Cliente do Hive-hbase.
hive-metastore-server	2.3.5-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.5-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.4.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	0.9.6	Servidor de vários usuários para blocos de anotações Jupyter

Componente	Version (Versão)	Descrição
livy-server	0.6.0-incubating	REST interface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.4.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.2.88	Drivers NVIDIA e toolkit CUDA
oozie-client	5.1.0	O cliente da linha de comando do Oozie.
oozie-server	5.1.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	3.4.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.2- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.2- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API

Componente	Version (Versão)	Descrição
presto-coordinator	0.220	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.220	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.1	O projeto R para computação estatística
spark-client	2.4.3	Os clientes da linha de comando do Spark.
spark-history-server	2.4.3	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.4.3	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.4.3	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	1.13.1	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.

Componente	Version (Versão)	Descrição
tez-on-yarn	0.9.2	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.8.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.14	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.14	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.26.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.26.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.

Classificações	Descrição
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.

Classificações	Descrição
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.

Classificações	Descrição
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
https-env	Mude os valores no HTTPFS ambiente.
https-site	Alterar os valores no arquivo https-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.

Classificações	Descrição
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.

Classificações	Descrição
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.

Classificações	Descrição
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.
recordserver-env	Mude os valores no EMR RecordServer ambiente.
recordserver-conf	Altere os valores no arquivo EMR RecordServer.properties do.
recordserver-log4j	Altere os valores no EMR RecordServer arquivo log4j.properties.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.

Classificações	Descrição
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.25.0

Versões 5.25.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.25.0	emr-5.24.1	emr-5.24.0	emr-5.23.1
AWS SDK para Java	1.11.566	1.11.546	1.11.546	1.11.519
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
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

Notas da versão 5.25.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.25.0 da Amazon. As alterações são referentes à versão 5.24.1.

Data da versão inicial: 17 de julho de 2019

Data da última atualização: 30 de outubro de 2019

Amazon EMR 5.25.0

Atualizações

- AWS SDK for Java 1.11.566
- Hive 2.3.5
- Presto 0.220
- Spark 2.4.3
- TensorFlow 1.13.1
- Tez 0.9.2
- Zookeeper 3.4.14

Novos atributos

- (30 de outubro de 2019) A partir da EMR versão 5.25.0 da Amazon, você pode se conectar à interface do usuário do servidor de histórico do Spark na página de resumo do cluster ou na guia Histórico do aplicativo no console. Em vez de configurar um proxy web por meio de uma SSH conexão, você pode acessar rapidamente a interface do servidor de histórico do Spark para visualizar as métricas do aplicativo e acessar arquivos de log relevantes para clusters ativos e

encerrados. Para obter mais informações, consulte [Acesso fora do cluster a interfaces de usuário de aplicativos persistentes](#) no Amazon EMR Management Guide.

Alterações, melhorias e problemas resolvidos

- Spark
 - Aprimorada a performance de algumas junções usando filtros de Bloom para pré-filtrar as entradas. A otimização é desabilitada por padrão e pode ser habilitada com a definição do parâmetro `spark.sql.bloomFilterJoin.enabled` de configuração do Spark como `true`.
 - Aprimorada a performance do agrupamento por colunas do tipo string.
 - Melhorou a memória padrão do executor Spark e a configuração dos núcleos dos tipos de instância R4 para clusters sem instalação. HBase
 - Resolvido um problema anterior com o atributo de remoção dinâmica de partições, em que a tabela removida precisava estar no lado esquerdo da junção.
 - Aprimorado DISTINCT antes da INTERSECT otimização para ser aplicado a casos adicionais envolvendo aliases.
 - Inferência aprimorada das estatísticas do SQL plano para DISTINCT consultas JOIN seguidas. Essa melhoria é desabilitada por padrão e pode ser habilitada pela definição do parâmetro `spark.sql.statsImprovements.enabled` de configuração do Spark como `true`. Essa otimização é exigida pelo atributo Distinct antes do Intersect e será habilitada automaticamente quando `spark.sql.optimizer.distinctBeforeIntersect.enabled` estiver definido como `true`.
 - Otimizada a ordem de junção com base no tamanho da tabela e nos filtros. Essa otimização é desativada por padrão e pode ser ativada com a definição do parâmetro `spark.sql.optimizer.sizeBasedJoinReorder.enabled` de configuração do Spark como `true`.

Para obter mais informações, consulte [Otimizar a performance do Spark](#).

- EMRFS
 - A EMRFS configuração, `fs.s3.buckets.create.enabled`, agora está desativada por padrão. Por meio de testes, descobrimos que a desabilitação dessa configuração melhora a performance e evita a criação não intencional de buckets do S3. Se sua aplicação depende dessa funcionalidade, você pode habilitá-la definindo a propriedade `fs.s3.buckets.create.enabled` como `true` na classificação de configuração `emrfs-site`. Para obter informações, consulte [Supplying a Configuration when Creating a Cluster](#).

- Melhorias na criptografia de disco local e na criptografia do S3 nas configurações de segurança (5 de agosto de 2019)
 - Separadas as configurações de criptografia do Amazon S3 das configurações de criptografia de disco local na configuração de segurança.
 - Foi adicionada uma opção para ativar a EBS criptografia com a versão 5.24.0 e posterior. Selecionar essa opção criptografa o volume do dispositivo raiz, além dos volumes de armazenamento. As versões anteriores exigiam o uso de um recurso personalizado AMI para criptografar o volume do dispositivo raiz.
 - Para obter mais informações, consulte [Opções de criptografia](#) no Amazon EMR Management Guide.

Problemas conhecidos

- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSH como hadoop usuário do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do hadoop.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo keytab está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas

definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

Versões de componente 5.25.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.2.4	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.10.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.9.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.11.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.

Componente	Version (Versão)	Descrição
emr-s3-select	1.3.0	EMRConector S3Select
emrfs	2.34.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.8.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.5-amzn-4	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-4	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.8.5-amzn-4	HDFScliente e biblioteca de linha de comando

Componente	Version (Versão)	Descrição
hadoop-hdfs-namenode	2.8.5-amzn-4	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	2.8.5-amzn-4	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httfs-server	2.8.5-amzn-4	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.8.5-amzn-4	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-4	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.8.5-amzn-4	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.8.5-amzn-4	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.5-amzn-4	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.9	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.

Componente	Version (Versão)	Descrição
hbase-region-server	1.4.9	Serviço para atender uma ou mais HBase regiões.
hbase-client	1.4.9	HBase cliente de linha de comando.
hbase-rest-server	1.4.9	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.4.9	Serviço que fornece um endpoint Thrift para HBase
hcatalog-client	2.3.5-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.5-amzn-0	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.5-amzn-0	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.3.5-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.3.5-amzn-0	Cliente do Hive-hbase.
hive-metastore-server	2.3.5-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL

Componente	Version (Versão)	Descrição
hive-server2	2.3.5-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.4.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	0.9.6	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.6.0-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.4.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.2.88	Drivers NVIDIA e toolkit CUDA
oozie-client	5.1.0	O cliente da linha de comando do Oozie.
oozie-server	5.1.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.

Componente	Version (Versão)	Descrição
opencv	3.4.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.1- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.1- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.220	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.220	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.1	O projeto R para computação estatística
spark-client	2.4.3	Os clientes da linha de comando do Spark.
spark-history-server	2.4.3	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.4.3	Mecanismo de execução na memória para YARN.

Componente	Version (Versão)	Descrição
spark-yarn-slave	2.4.3	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	1.13.1	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.2	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.8.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.14	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.14	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.25.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.25.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.
container-log4j	Altere os valores no arquivo container-log4j.properties YARN do Hadoop.
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4j-yarn-session.properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase

Classificações	Descrição
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.

Classificações	Descrição
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.

Classificações	Descrição
jupyter-notebook-conf	Alterar os valores no arquivo <code>jupyter_notebook_config.py</code> do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo <code>jupyterhub_config.py</code> do.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.
jupyter-sparkmagic-conf	Altere os valores no arquivo <code>config.json</code> do Sparkmagic.
livy-conf	Alterar os valores no arquivo <code>livy.conf</code> do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de <code>log4j.properties</code> no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo <code>mapred-site.xml</code> do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo <code>oozie-log4j.properties</code> do Oozie.
oozie-site	Alterar os valores no arquivo <code>oozie-site.xml</code> do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo <code>hadoop-metrics2-hbase.properties</code> do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo <code>hbase-site.xml</code> do Phoenix.

Classificações	Descrição
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.

Classificações	Descrição
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.
recordserver-env	Mude os valores no EMR RecordServer ambiente.

Classificações	Descrição
recordserver-conf	Altere os valores no arquivo EMR RecordServer <code>erver.properties</code> do.
recordserver-log4j	Altere os valores no EMR RecordServer arquivo <code>log4j.properties</code> .
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo <code>spark-defaults.conf</code> do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo <code>hive-site.xml</code> do Spark.
spark-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Spark.
spark-metrics	Alterar os valores no arquivo <code>metrics.properties</code> do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo <code>oraoop-site.xml</code> OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo <code>sqoop-site.xml</code> do Sqoop.
tez-site	Alterar os valores no arquivo <code>tez-site.xml</code> do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo <code>yarn-site.xml</code> do.

Classificações	Descrição
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.24.1

Versões 5.24.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.24.1	emr-5.24.0	emr-5.23.1	emr-5.23.0
AWS SDK para Java	1.11.546	1.11.546	1.11.519	1.11.519
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6

	emr-5.24.1	emr-5.24.0	emr-5.23.1	emr-5.23.0
Scala	2.11.12	2.11.12	2.11.12	2.11.12
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
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

	emr-5.24.1	emr-5.24.0	emr-5.23.1	emr-5.23.0
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
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

Notas da versão 5.24.1

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.24.1 da Amazon. As alterações são referentes à versão 5.24.0.

Data da versão inicial: 26 de junho de 2019

Alterações, melhorias e problemas resolvidos

- Atualizou o Amazon Linux padrão AMI para Amazon EMR para incluir atualizações importantes de segurança do kernel Linux, incluindo o problema de TCP SACK negação de serviço ([AWS-2019-005](#)).

Problemas conhecidos

- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como

redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSH como usuário do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do hadoop.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo keytab está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

Versões de componente 5.24.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
aws-sagemaker-spark-sdk	1.2.1	Amazon SageMaker Spark SDK
emr-ddb	4.9.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.8.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.11.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	1.3.0	EMRConector S3Select
emrfs	2.33.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.8.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.

Componente	Version (Versão)	Descrição
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.5-amzn-4	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-4	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.8.5-amzn-4	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.8.5-amzn-4	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	2.8.5-amzn-4	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httpfs-server	2.8.5-amzn-4	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.8.5-amzn-4	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-4	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.

Componente	Version (Versão)	Descrição
hadoop-yarn-nodemanager	2.8.5-amzn-4	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.8.5-amzn-4	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.5-amzn-4	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.9	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.9	Serviço para atender uma ou mais HBase regiões.
hbase-client	1.4.9	HBasecliente de linha de comando.
hbase-rest-server	1.4.9	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	1.4.9	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	2.3.4-amzn-2	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.

Componente	Version (Versão)	Descrição
hcatalog-server	2.3.4-amzn-2	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.4-amzn-2	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	2.3.4-amzn-2	O cliente da linha de comando do Hive.
hive-hbase	2.3.4-amzn-2	Cliente do Hive-hbase.
hive-metastore-server	2.3.4-amzn-2	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.4-amzn-2	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.4.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	0.9.6	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.6.0-incubating	RESTinterface para interagir com o Apache Spark

Componente	Version (Versão)	Descrição
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.4.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.2.88	Drivers NVIDIA e toolkit CUDA
oozie-client	5.1.0	O cliente da linha de comando do Oozie.
oozie-server	5.1.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	3.4.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.1- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.1- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.219	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.

Componente	Version (Versão)	Descrição
presto-worker	0.219	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.1	O projeto R para computação estatística
spark-client	2.4.2	Os clientes da linha de comando do Spark.
spark-history-server	2.4.2	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.4.2	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.4.2	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	1.12.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.1	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.

Componente	Version (Versão)	Descrição
zeppelin-server	0.8.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.13	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.13	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.24.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.24.1

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .

Classificações	Descrição
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.

Classificações	Descrição
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.

Classificações	Descrição
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.

Classificações	Descrição
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.

Classificações	Descrição
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.

Classificações	Descrição
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.

Classificações	Descrição
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.24.0

Versões 5.24.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.24.0	emr-5.23.1	emr-5.23.0	emr-5.22.0
AWS SDK para Java	1.11.546	1.11.519	1.11.519	1.11.510
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
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	-	-	-	-
Hue	4.4.0	4.3.0	4.3.0	4.3.0
Iceberg	-	-	-	-

	emr-5.24.0	emr-5.23.1	emr-5.23.0	emr-5.22.0
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

Notas da versão 5.24.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.24.0 da Amazon. As alterações são referentes à versão 5.23.0.

Data da versão inicial: 11 de junho de 2019

Data da última atualização: 5 de agosto de 2019

Atualizações

- 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 1.11.546
- Conectores e drivers:
 - Conector do DynamoDB 4.9.0
 - Conector do MariaDB 2.4.1
 - Driver Amazon Redshift 1.2.27.1051 JDBC

Alterações, melhorias e problemas resolvidos

- Spark
 - Adicionada otimização para remover partições dinamicamente. A otimização está desabilitada por padrão. Para habilitá-la, defina o parâmetro `spark.sql.dynamicPartitionPruning.enabled` de configuração do Spark como `true`.
 - Aprimorada a performance de consultas INTERSECT. Essa otimização está desabilitada por padrão. Para habilitá-la, defina o parâmetro `spark.sql.optimizer.distinctBeforeIntersect.enabled` de configuração do Spark como `true`.
 - Adicionada otimização para nivelar subconsultas escalares com agregados que usam a mesma relação. A otimização está desabilitada por padrão. Para habilitá-la, defina o parâmetro `spark.sql.optimizer.flattenScalarSubqueriesWithAggregates.enabled` de configuração do Spark como `true`.
 - Aprimorada a geração de código em todo o estágio.

Para obter mais informações, consulte [Otimizar a performance do Spark](#).

- Melhorias na criptografia de disco local e na criptografia do S3 nas configurações de segurança (5 de agosto de 2019)
 - Separadas as configurações de criptografia do Amazon S3 das configurações de criptografia de disco local na configuração de segurança.
 - Foi adicionada uma opção para ativar a EBS criptografia. Selecionar essa opção criptografa o volume do dispositivo raiz, além dos volumes de armazenamento. As versões anteriores exigiam o uso de um recurso personalizado AMI para criptografar o volume do dispositivo raiz.
 - Para obter mais informações, consulte [Opções de criptografia](#) no Amazon EMR Management Guide.

Problemas conhecidos

- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSH como hadoop usuário do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do hadoop.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo keytab está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas

definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

Versões de componente 5.24.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.2.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.9.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.8.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.11.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.

Componente	Version (Versão)	Descrição
emr-s3-select	1.3.0	EMRConector S3Select
emrfs	2.33.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.8.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.5-amzn-4	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-4	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.8.5-amzn-4	HDFScliente e biblioteca de linha de comando

Componente	Version (Versão)	Descrição
hadoop-hdfs-namenode	2.8.5-amzn-4	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	2.8.5-amzn-4	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httpfs-server	2.8.5-amzn-4	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.8.5-amzn-4	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-4	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.8.5-amzn-4	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.8.5-amzn-4	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.5-amzn-4	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.9	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.

Componente	Version (Versão)	Descrição
hbase-region-server	1.4.9	Serviço para atender uma ou mais HBase regiões.
hbase-client	1.4.9	HBase cliente de linha de comando.
hbase-rest-server	1.4.9	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.4.9	Serviço que fornece um endpoint Thrift para HBase
hcatalog-client	2.3.4-amzn-2	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.4-amzn-2	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.4-amzn-2	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.3.4-amzn-2	O cliente da linha de comando do Hive.
hive-hbase	2.3.4-amzn-2	Cliente do Hive-hbase.
hive-metastore-server	2.3.4-amzn-2	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL

Componente	Version (Versão)	Descrição
hive-server2	2.3.4-amzn-2	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.4.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	0.9.6	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.6.0-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.4.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.2.88	Drivers NVIDIA e toolkit CUDA
oozie-client	5.1.0	O cliente da linha de comando do Oozie.
oozie-server	5.1.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.

Componente	Version (Versão)	Descrição
opencv	3.4.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.1- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.1- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.219	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.219	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.1	O projeto R para computação estatística
spark-client	2.4.2	Os clientes da linha de comando do Spark.
spark-history-server	2.4.2	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.4.2	Mecanismo de execução na memória para YARN.

Componente	Version (Versão)	Descrição
spark-yarn-slave	2.4.2	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	1.12.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.1	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.8.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.13	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.13	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.24.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.24.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.
container-log4j	Altere os valores no arquivo container-log4j.properties YARN do Hadoop.
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4j-yarn-session.properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase

Classificações	Descrição
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.

Classificações	Descrição
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.

Classificações	Descrição
jupyter-notebook-conf	Alterar os valores no arquivo <code>jupyter_notebook_config.py</code> do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo <code>jupyterhub_config.py</code> do.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.
jupyter-sparkmagic-conf	Altere os valores no arquivo <code>config.json</code> do Sparkmagic.
livy-conf	Alterar os valores no arquivo <code>livy.conf</code> do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de <code>log4j.properties</code> no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo <code>mapred-site.xml</code> do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo <code>oozie-log4j.properties</code> do Oozie.
oozie-site	Alterar os valores no arquivo <code>oozie-site.xml</code> do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo <code>hadoop-metrics2-hbase.properties</code> do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo <code>hbase-site.xml</code> do Phoenix.

Classificações	Descrição
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.

Classificações	Descrição
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.

Classificações	Descrição
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.23.1

Versões 5.23.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livvy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.23.1	emr-5.23.0	emr-5.22.0	emr-5.21.2
AWS SDK para Java	1.11.519	1.11.519	1.11.510	1.11.479
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
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

	emr-5.23.1	emr-5.23.0	emr-5.22.0	emr-5.21.2
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	5.1.0	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

	emr-5.23.1	emr-5.23.0	emr-5.22.0	emr-5.21.2
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

Notas da versão 5.23.1

Esta é uma versão de patch. Todos os aplicativos e componentes são iguais aos da EMR versão anterior da Amazon.

Status de suporte do Instance Metadata Service (IMDS) V2: os componentes Amazon EMR 5.23.1, 5.27.1 e 5.32 ou posteriores são usados para todas as chamadas. IMDSv2 IMDS Para IMDS chamadas no código do seu aplicativo, você pode usar ambos IMDSv1 e IMDSv2, ou configurar o IMDS para uso somente IMDSv2 para aumentar a segurança. Para outras EMR versões 5.x, a desativação IMDSv1 causa falha na inicialização do cluster.

Versões de componente 5.23.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
aws-sagemaker-spark-sdk	1.2.1	Amazon SageMaker Spark SDK
emr-ddb	4.8.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.7.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.11.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	1.2.0	EMRConector S3Select
emrfs	2.32.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.7.1	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.

Componente	Version (Versão)	Descrição
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.5-amzn-3	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-3	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.8.5-amzn-3	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.8.5-amzn-3	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	2.8.5-amzn-3	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.
hadoop-httpfs-server	2.8.5-amzn-3	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.8.5-amzn-3	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-3	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.

Componente	Version (Versão)	Descrição
hadoop-yarn-nodemanager	2.8.5-amzn-3	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.8.5-amzn-3	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.5-amzn-3	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.9	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.9	Serviço para atender uma ou mais HBase regiões.
hbase-client	1.4.9	HBasecliente de linha de comando.
hbase-rest-server	1.4.9	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	1.4.9	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	2.3.4-amzn-1	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.

Componente	Version (Versão)	Descrição
hcatalog-server	2.3.4-amzn-1	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.4-amzn-1	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	2.3.4-amzn-1	O cliente da linha de comando do Hive.
hive-hbase	2.3.4-amzn-1	Cliente do Hive-hbase.
hive-metastore-server	2.3.4-amzn-1	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.4-amzn-1	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.3.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	0.9.4	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.5.0-incubating	RESTinterface para interagir com o Apache Spark

Componente	Version (Versão)	Descrição
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.3.1	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.2.88	Drivers NVIDIA e toolkit CUDA
oozie-client	5.1.0	O cliente da linha de comando do Oozie.
oozie-server	5.1.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	3.4.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.1- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.1- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.215	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.

Componente	Version (Versão)	Descrição
presto-worker	0.215	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.1	O projeto R para computação estatística
spark-client	2.4.0	Os clientes da linha de comando do Spark.
spark-history-server	2.4.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.4.0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.4.0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	1.12.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.1	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.

Componente	Version (Versão)	Descrição
zeppelin-server	0.8.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.13	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.13	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.23.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.23.1

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .

Classificações	Descrição
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.

Classificações	Descrição
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.

Classificações	Descrição
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.

Classificações	Descrição
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.

Classificações	Descrição
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.

Classificações	Descrição
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.

Classificações	Descrição
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.23.0

Versões 5.23.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.23.0	emr-5.22.0	emr-5.21.2	emr-5.21.1
AWS SDK para Java	1.11.519	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	2.11.12	2.11.12	2.11.12	2.11.12
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	-	-	-	-

	emr-5.23.0	emr-5.22.0	emr-5.21.2	emr-5.21.1
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	5.0.0	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

Notas da versão 5.23.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.23.0 da Amazon. As alterações são referentes à versão 5.22.0.

Data da versão inicial: 1.º de abril de 2019

Data da última atualização: 30 de abril de 2019

Atualizações

- AWS SDK for Java 1.11.519

Novos atributos

- (30 de abril de 2019) Com o Amazon EMR 5.23.0 e versões posteriores, você pode iniciar um cluster com três nós principais para oferecer suporte à alta disponibilidade de aplicativos como YARN Resource Manager, Spark HDFS NameNode, Hive e Ganglia. O nó primário não é mais um possível ponto de falha único com esse recurso. Se um dos nós primários falhar, a Amazon EMR automaticamente passa para um nó primário em espera e substitui o nó primário com falha por um novo com as mesmas ações de configuração e bootstrap. Para obter mais informações, consulte [Plan and Configure Primary Nodes](#).

Problemas conhecidos

- Tez UI (corrigido na EMR versão 5.26.0 da Amazon)

O Tez UI não funciona em um EMR cluster com vários nós primários.

- Hue (corrigido na EMR versão 5.24.0 da Amazon)
 - O Hue executado na Amazon EMR não oferece suporte ao Solr. A partir da EMR versão 5.20.0 da Amazon, um problema de configuração incorreta faz com que o Solr seja ativado e uma mensagem de erro inofensiva apareça semelhante à seguinte:

```
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',))
```

Para evitar que a mensagem de erro do Solr seja exibida:

1. Conecte-se à linha de comando do nó primário usando SSH o.
2. Use um editor de texto para abrir o arquivo `hue.ini`. Por exemplo: .

```
sudo vim /etc/hue/conf/hue.ini
```

3. Pesquise o termo `appblacklist` e modifique a linha para o seguinte:

```
appblacklist = search
```

4. Salve as alterações e reinicie o Hue, conforme mostrado no exemplo a seguir:

```
sudo stop hue; sudo start hue
```

- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSH como `hadoop` usuário do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do `hadoop`.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo `keytab` está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

Versões de componente 5.23.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.2.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.8.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.7.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.11.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
<code>emr-s3-select</code>	1.2.0	EMRConector S3Select
<code>emrfs</code>	2.32.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
flink-client	1.7.1	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.5-amzn-3	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-3	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.8.5-amzn-3	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.8.5-amzn-3	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-hdfs-journalnode	2.8.5-amzn-3	HDFSserviço para gerenciar o diário do sistema de arquivos Hadoop em clusters HA.

Componente	Version (Versão)	Descrição
hadoop-https-server	2.8.5-amzn-3	HTTP endpoint para HDFS operações.
hadoop-kms-server	2.8.5-amzn-3	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-3	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.8.5-amzn-3	YARN serviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.8.5-amzn-3	YARN serviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.5-amzn-3	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.9	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.9	Serviço para atender uma ou mais HBase regiões.
hbase-client	1.4.9	HBase cliente de linha de comando.

Componente	Version (Versão)	Descrição
hbase-rest-server	1.4.9	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.4.9	Serviço que fornece um endpoint Thrift para HBase
hcatalog-client	2.3.4-amzn-1	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.4-amzn-1	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.4-amzn-1	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.3.4-amzn-1	O cliente da linha de comando do Hive.
hive-hbase	2.3.4-amzn-1	Cliente do Hive-hbase.
hive-metastore-server	2.3.4-amzn-1	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.4-amzn-1	O serviço que aceita as consultas do Hive como solicitações da web.

Componente	Version (Versão)	Descrição
hue-server	4.3.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	0.9.4	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.5.0-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.3.1	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.2.88	Drivers NVIDIA e toolkit CUDA
oozie-client	5.1.0	O cliente da linha de comando do Oozie.
oozie-server	5.1.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	3.4.0	Biblioteca de Visão Computacional de Código Aberto.

Componente	Version (Versão)	Descrição
phoenix-library	4.14.1- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.1- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.215	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.215	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.1	O projeto R para computação estatística
spark-client	2.4.0	Os clientes da linha de comando do Spark.
spark-history-server	2.4.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.4.0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.4.0	Bibliotecas do Apache Spark necessárias para escravos. YARN

Componente	Version (Versão)	Descrição
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	1.12.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.1	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.8.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.13	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.13	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.23.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.23.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.
container-log4j	Altere os valores no arquivo container-log4j.properties YARN do Hadoop.
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4j-yarn-session.properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase

Classificações	Descrição
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure as zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.

Classificações	Descrição
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.

Classificações	Descrição
jupyter-notebook-conf	Alterar os valores no arquivo <code>jupyter_notebook_config.py</code> do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo <code>jupyterhub_config.py</code> do.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.
jupyter-sparkmagic-conf	Altere os valores no arquivo <code>config.json</code> do Sparkmagic.
livy-conf	Alterar os valores no arquivo <code>livy.conf</code> do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de <code>log4j.properties</code> no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo <code>mapred-site.xml</code> do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo <code>oozie-log4j.properties</code> do Oozie.
oozie-site	Alterar os valores no arquivo <code>oozie-site.xml</code> do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo <code>hadoop-metrics2-hbase.properties</code> do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo <code>hbase-site.xml</code> do Phoenix.

Classificações	Descrição
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.

Classificações	Descrição
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.

Classificações	Descrição
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.22.0

Versões 5.22.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.22.0	emr-5.21.2	emr-5.21.1	emr-5.21.0
AWS SDK para 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
Scala	2.11.12	2.11.12	2.11.12	2.11.12
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

	emr-5.22.0	emr-5.21.2	emr-5.21.1	emr-5.21.0
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	4.3.0	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
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

	emr-5.22.0	emr-5.21.2	emr-5.21.1	emr-5.21.0
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

Notas da versão 5.22.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.22.0 da Amazon. As alterações são referentes à versão 5.21.0.

Important

A partir da EMR versão 5.22.0 da Amazon, a Amazon EMR usa o AWS Signature versão 4 exclusivamente para autenticar solicitações para o Amazon S3. As EMR versões anteriores da Amazon usam a AWS Signature Version 2 em alguns casos, a menos que as notas de lançamento indiquem que a Signature Version 4 é usada exclusivamente. Para obter mais informações, consulte [Autenticação de solicitações \(AWS assinatura versão 4\)](#) e [Solicitações de autenticação \(AWS assinatura versão 2\) no Guia](#) do desenvolvedor do Amazon Simple Storage Service.

Data da versão inicial: 20 de março de 2019

Atualizações

- Flink 1.7.1
- HBase1.4.9
- Oozie 5.1.0
- Phoenix 4.14.1
- Zeppelin 0.8.1
- Conectores e drivers:
 - Conector do DynamoDB 4.8.0
 - Conector do MariaDB 2.2.6

- Driver Amazon Redshift 1.2.20.1043 JDBC

Novos atributos

- Modificou a EBS configuração padrão para tipos de EC2 instância com armazenamento EBS somente. Quando você cria um cluster usando a EMR versão 5.22.0 e posterior da Amazon, a quantidade padrão de EBS armazenamento aumenta com base no tamanho da instância. Além disso, dividimos o aumento do armazenamento em vários volumes, proporcionando maior IOPS desempenho. Se quiser usar uma configuração de armazenamento de EBS instância diferente, você pode especificá-la ao criar um EMR cluster ou adicionar nós a um cluster existente. Para obter mais informações sobre a quantidade de armazenamento e o número de volumes alocados por padrão para cada tipo de instância, consulte [EBSArmazenamento padrão para instâncias](#) no Amazon EMR Management Guide.

Alterações, melhorias e problemas resolvidos

- Spark
 - Introduziu uma nova propriedade de configuração para o Spark em YARN, `spark.yarn.executor.memoryOverheadFactor`. O valor dessa propriedade é um fator de escala que define o valor da sobrecarga de memória como uma porcentagem da memória do executor, com um mínimo de 384 MB. Se a sobrecarga de memória for definida explicitamente usando `spark.yarn.executor.memoryOverhead`, essa propriedade não terá efeito. O valor padrão é `0.1875`, representando 18,75%. Esse padrão para a Amazon EMR deixa mais espaço nos YARN contêineres para sobrecarga de memória do executor do que o padrão de 10% definido internamente pelo Spark. O EMR padrão da Amazon de 18,75% mostrou empiricamente menos falhas relacionadas à memória nos benchmarks -DS. TPC
 - Backport [SPARK-26316](#) para melhorar o desempenho.
- Nas EMR versões 5.19.0, 5.20.0 e 5.21.0 da Amazon, os rótulos dos YARN nós são armazenados em um diretório. HDFS Em algumas situações, isso leva a atrasos na inicialização do nó central causando, em seguida, tempo limite do cluster e falha na inicialização. A partir do Amazon EMR 5.22.0, esse problema foi resolvido. YARN os rótulos dos nós são armazenados no disco local de cada nó do cluster, evitando dependências em HDFS.

Problemas conhecidos

- Hue (corrigido na EMR versão 5.24.0 da Amazon)

- O Hue executado na Amazon EMR não oferece suporte ao Solr. A partir da EMR versão 5.20.0 da Amazon, um problema de configuração incorreta faz com que o Solr seja ativado e uma mensagem de erro inofensiva apareça semelhante à seguinte:

```
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',))
```

Para evitar que a mensagem de erro do Solr seja exibida:

1. Conecte-se à linha de comando do nó primário usando SSH o.
2. Use um editor de texto para abrir o arquivo `hue.ini`. Por exemplo: .

```
sudo vim /etc/hue/conf/hue.ini
```

3. Pesquise o termo `appblacklist` e modifique a linha para o seguinte:

```
appblacklist = search
```

4. Salve as alterações e reinicie o Hue, conforme mostrado no exemplo a seguir:

```
sudo stop hue; sudo start hue
```

- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSH como `hadoop` usuário do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do `hadoop`.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo keytab está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

Versões de componente 5.22.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.2.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.8.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
emr-goodies	2.6.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.11.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	1.2.0	EMRConector S3Select
emrfs	2.31.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.7.1	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.

Componente	Version (Versão)	Descrição
hadoop-client	2.8.5-amzn-2	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-2	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.8.5-amzn-2	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.8.5-amzn-2	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-https-server	2.8.5-amzn-2	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.8.5-amzn-2	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-2	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.8.5-amzn-2	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.8.5-amzn-2	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.5-amzn-2	Serviço para recuperar informações atuais e históricas para YARN aplicativos.

Componente	Version (Versão)	Descrição
hbase-hmaster	1.4.9	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.9	Serviço para atender uma ou mais HBase regiões.
hbase-client	1.4.9	HBase cliente de linha de comando.
hbase-rest-server	1.4.9	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.4.9	Serviço que fornece um endpoint Thrift para HBase.
hcatalog-client	2.3.4-amzn-1	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.4-amzn-1	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.4-amzn-1	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.3.4-amzn-1	O cliente da linha de comando do Hive.
hive-hbase	2.3.4-amzn-1	Cliente do Hive-hbase.

Componente	Version (Versão)	Descrição
hive-metastore-server	2.3.4-amzn-1	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.4-amzn-1	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.3.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	0.9.4	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.5.0-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.3.1	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.2.88	Drivers NVIDIA e toolkit CUDA

Componente	Version (Versão)	Descrição
oozie-client	5.1.0	O cliente da linha de comando do Oozie.
oozie-server	5.1.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	3.4.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.1- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.1- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.215	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.215	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.1	O projeto R para computação estatística
spark-client	2.4.0	Os clientes da linha de comando do Spark.

Componente	Version (Versão)	Descrição
spark-history-server	2.4.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.4.0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.4.0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	1.12.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.1	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.
zeppelin-server	0.8.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.13	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.

Componente	Version (Versão)	Descrição
zookeeper-client	3.4.13	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.22.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.22.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Alterar as configurações de <code>log4j-cli.properties</code> no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.

Classificações	Descrição
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure as zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.

Classificações	Descrição
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.

Classificações	Descrição
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.

Classificações	Descrição
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.

Classificações	Descrição
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.

Classificações	Descrição
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.

Classificações	Descrição
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.21.2

Versões 5.21.2 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.21.2	emr-5.21.1	emr-5.21.0	emr-5.20.1
AWS SDK para Java	1.11.479	1.11.479	1.11.479	1.11.461

	emr-5.21.2	emr-5.21.1	emr-5.21.0	emr-5.20.1
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,6
Scala	2.11.12	2.11.12	2.11.12	2.11.12
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
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

	emr-5.21.2	emr-5.21.1	emr-5.21.0	emr-5.20.1
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
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

Notas da versão 5.21.2

Esta é uma versão de patch para adicionar a autenticação AWS Signature versão 4 para solicitações ao Amazon S3. Todos os aplicativos e componentes são iguais aos da EMR versão anterior da Amazon.

Important

Nesta versão de lançamento, a Amazon EMR usa o AWS Signature versão 4 exclusivamente para autenticar solicitações para o Amazon S3. Para obter mais informações, consulte [Novidades](#).

Versões de componente 5.21.2

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR

e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.2.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.7.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.5.1	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.11.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
<code>emr-s3-select</code>	1.2.0	EMRConector S3Select
<code>emrfs</code>	2.30.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
<code>flink-client</code>	1.7.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.

Componente	Version (Versão)	Descrição
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.5-amzn-1	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.8.5-amzn-1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.8.5-amzn-1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httpfs-server	2.8.5-amzn-1	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.8.5-amzn-1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API

Componente	Version (Versão)	Descrição
hadoop-mapred	2.8.5-amzn-1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.8.5-amzn-1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.8.5-amzn-1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.5-amzn-1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.8	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.8	Serviço para atender uma ou mais HBase regiões.
hbase-client	1.4.8	HBasecliente de linha de comando.
hbase-rest-server	1.4.8	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	1.4.8	Serviço que fornece um endpoint Thrift para. HBase

Componente	Version (Versão)	Descrição
hcatalog-client	2.3.4-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.4-amzn-0	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.4-amzn-0	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	2.3.4-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.3.4-amzn-0	Cliente do Hive-hbase.
hive-metastore-server	2.3.4-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.4-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.3.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	0.9.4	Servidor de vários usuários para blocos de anotações Jupyter

Componente	Version (Versão)	Descrição
livy-server	0.5.0-incubating	REST interface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.3.1	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.2.88	Drivers NVIDIA e toolkit CUDA
oozie-client	5.0.0	O cliente da linha de comando do Oozie.
oozie-server	5.0.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	3.4.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.0- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.0- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API

Componente	Version (Versão)	Descrição
presto-coordinator	0.215	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.215	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.1	O projeto R para computação estatística
spark-client	2.4.0	Os clientes da linha de comando do Spark.
spark-history-server	2.4.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.4.0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.4.0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	1.12.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.

Componente	Version (Versão)	Descrição
tez-on-yarn	0.9.1	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.8.0	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.13	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.13	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.21.2

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.21.2

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.

Classificações	Descrição
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.

Classificações	Descrição
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure as zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.

Classificações	Descrição
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
https-env	Mude os valores no HTTPFS ambiente.
https-site	Alterar os valores no arquivo https-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.

Classificações	Descrição
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.

Classificações	Descrição
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.

Classificações	Descrição
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.

Classificações	Descrição
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.21.1

Versões 5.21.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.21.1	emr-5.21.0	emr-5.20.1	emr-5.20.0
AWS SDK para 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	2.11.12	2.11.12	2.11.12	2.11.12
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

	emr-5.21.1	emr-5.21.0	emr-5.20.1	emr-5.20.0
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	0.214	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

Notas da versão 5.21.1

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.21.1 da Amazon. As alterações são referentes à versão 5.21.0.

Data da versão inicial: 18 de julho de 2019

Alterações, melhorias e problemas resolvidos

- Atualizou o Amazon Linux padrão AMI para Amazon EMR para incluir atualizações importantes de segurança do kernel Linux, incluindo o problema de TCP SACK negação de serviço ([AWS-2019-005](#)).

Problemas conhecidos

- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSH como hadoop usuário do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do hadoop.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo keytab está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas

definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

Versões de componente 5.21.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR diferem das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.2.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.7.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.5.1	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.11.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.

Componente	Version (Versão)	Descrição
emr-s3-select	1.2.0	EMRConector S3Select
emrfs	2.30.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.7.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.5-amzn-1	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.8.5-amzn-1	HDFScliente e biblioteca de linha de comando

Componente	Version (Versão)	Descrição
hadoop-hdfs-namenode	2.8.5-amzn-1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httfs-server	2.8.5-amzn-1	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.8.5-amzn-1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.8.5-amzn-1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.8.5-amzn-1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.5-amzn-1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.8	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.8	Serviço para atender a uma ou mais HBase regiões.

Componente	Version (Versão)	Descrição
hbase-client	1.4.8	HBase cliente de linha de comando.
hbase-rest-server	1.4.8	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.4.8	Serviço que fornece um endpoint Thrift para HBase
hcatalog-client	2.3.4-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.4-amzn-0	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.4-amzn-0	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.3.4-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.3.4-amzn-0	Cliente do Hive-hbase.
hive-metastore-server	2.3.4-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL

Componente	Version (Versão)	Descrição
hive-server2	2.3.4-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.3.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	0.9.4	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.5.0-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.3.1	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.2.88	Drivers NVIDIA e toolkit CUDA
oozie-client	5.0.0	O cliente da linha de comando do Oozie.
oozie-server	5.0.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.

Componente	Version (Versão)	Descrição
opencv	3.4.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.0- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.0- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.215	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.215	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.1	O projeto R para computação estatística
spark-client	2.4.0	Os clientes da linha de comando do Spark.
spark-history-server	2.4.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.4.0	Mecanismo de execução na memória para YARN.

Componente	Version (Versão)	Descrição
spark-yarn-slave	2.4.0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	1.12.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.1	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.8.0	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.13	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.13	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.21.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.21.1

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.
container-log4j	Altere os valores no arquivo container-log4j.properties YARN do Hadoop.
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4j-yarn-session.properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase

Classificações	Descrição
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.

Classificações	Descrição
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente do HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.

Classificações	Descrição
jupyter-notebook-conf	Alterar os valores no arquivo <code>jupyter_notebook_config.py</code> do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo <code>jupyterhub_config.py</code> do.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.
jupyter-sparkmagic-conf	Altere os valores no arquivo <code>config.json</code> do Sparkmagic.
livy-conf	Alterar os valores no arquivo <code>livy.conf</code> do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de <code>log4j.properties</code> no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo <code>mapred-site.xml</code> do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo <code>oozie-log4j.properties</code> do Oozie.
oozie-site	Alterar os valores no arquivo <code>oozie-site.xml</code> do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo <code>hadoop-metrics2-hbase.properties</code> do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo <code>hbase-site.xml</code> do Phoenix.

Classificações	Descrição
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.

Classificações	Descrição
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.

Classificações	Descrição
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.21.0

Versões 5.21.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.21.0	emr-5.20.1	emr-5.20.0	emr-5.19.1
AWS SDK para Java	1.11.479	1.11.461	1.11.461	1.11.433
Python	2,7, 3,6	2,7, 3,6	2,7, 3,6	2,7, 3,4
Scala	2.11.12	2.11.12	2.11.12	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

	emr-5.21.0	emr-5.20.1	emr-5.20.0	emr-5.19.1
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
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

	emr-5.21.0	emr-5.20.1	emr-5.20.0	emr-5.19.1
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

Notas da versão 5.21.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.21.0 da Amazon. As alterações são referentes à versão 5.20.0.

Data da versão inicial: 18 de fevereiro de 2019

Data da última atualização: 3 de abril de 2019

Atualizações

- Flink 1.7.0
- Presto 0.215
- AWS SDK for Java 1.11.479

Novos atributos

- (3 de abril de 2019) Com a Amazon EMR versão 5.21.0 e posterior, você pode substituir as configurações do cluster e especificar classificações de configuração adicionais para cada grupo de instâncias em um cluster em execução. Você faz isso usando o EMR console da Amazon, o AWS Command Line Interface (AWS CLI) ou AWS SDK o. Para obter mais informações, consulte [Supplying a Configuration for an Instance Group in a Running Cluster](#).

Alterações, melhorias e problemas resolvidos

- Zeppelin
 - Porta traseira [ZEPPELIN-3878](#).

Problemas conhecidos

- Hue (corrigido na EMR versão 5.24.0 da Amazon)
 - O Hue executado na Amazon EMR não oferece suporte ao Solr. A partir da EMR versão 5.20.0 da Amazon, um problema de configuração incorreta faz com que o Solr seja ativado e uma mensagem de erro inofensiva apareça semelhante à seguinte:

```
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',))
```

Para evitar que a mensagem de erro do Solr seja exibida:

1. Conecte-se à linha de comando do nó primário usando SSH o.
2. Use um editor de texto para abrir o arquivo `hue.ini`. Por exemplo: .

```
sudo vim /etc/hue/conf/hue.ini
```

3. Pesquise o termo `appblacklist` e modifique a linha para o seguinte:

```
appblacklist = search
```

4. Salve as alterações e reinicie o Hue, conforme mostrado no exemplo a seguir:

```
sudo stop hue; sudo start hue
```

- Tez
 - Esse problema foi corrigido no Amazon EMR 5.22.0.

Quando você se conecta à interface do usuário do Tez em `http://MasterDNS:8080/tez-ui` por meio de uma SSH conexão com o nó primário do cluster, surge o erro “Falha na operação do adaptador - Timeline server () está fora de alcance. ATS “Está inativo ou não CORS está ativado” aparece ou as tarefas mostram inesperadamente N/A.

Isso é causado pela interface do usuário do Tez fazendo solicitações ao YARN Timeline Server usando, `localhost` em vez do nome do host, o nó primário. Como solução alternativa, um script está disponível para execução como ação ou etapa de bootstrap. O script atualiza o nome

do host no arquivo `configs.env` do Tez. Para obter mais informações e a localização do script, consulte [Instruções de bootstrap](#).

- Nas EMR versões 5.19.0, 5.20.0 e 5.21.0 da Amazon, os rótulos dos YARN nós são armazenados em um diretório. HDFS Em algumas situações, isso leva a atrasos na inicialização do nó central causando, em seguida, tempo limite do cluster e falha na inicialização. A partir do Amazon EMR 5.22.0, esse problema foi resolvido. YARNos rótulos dos nós são armazenados no disco local de cada nó do cluster, evitando dependências emHDFS.
- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSHcomo hadoop usuário do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do hadoop.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo keytab está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

Versões de componente 5.21.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR diferem das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.2.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.7.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.5.1	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.11.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
<code>emr-s3-select</code>	1.2.0	EMRConector S3Select
<code>emrfs</code>	2.30.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
flink-client	1.7.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.5-amzn-1	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.8.5-amzn-1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.8.5-amzn-1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-https-server	2.8.5-amzn-1	HTTPendpoint para HDFS operações.

Componente	Version (Versão)	Descrição
hadoop-kms-server	2.8.5-amzn-1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.8.5-amzn-1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.8.5-amzn-1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.5-amzn-1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.8	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.8	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.4.8	HBasecliente de linha de comando.
hbase-rest-server	1.4.8	Serviço que fornece um RESTful HTTP endpoint paraHBase.

Componente	Version (Versão)	Descrição
hbase-thrift-server	1.4.8	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	2.3.4-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.4-amzn-0	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.4-amzn-0	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	2.3.4-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.3.4-amzn-0	Cliente do Hive-hbase.
hive-metastore-server	2.3.4-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.4-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.3.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop

Componente	Version (Versão)	Descrição
jupyterhub	0.9.4	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.5.0-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.3.1	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.2.88	Drivers NVIDIA e toolkit CUDA
oozie-client	5.0.0	O cliente da linha de comando do Oozie.
oozie-server	5.0.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	3.4.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.0- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente

Componente	Version (Versão)	Descrição
phoenix-query-server	4.14.0- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.215	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.215	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.1	O projeto R para computação estatística
spark-client	2.4.0	Os clientes da linha de comando do Spark.
spark-history-server	2.4.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.4.0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.4.0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.

Componente	Version (Versão)	Descrição
tensorflow	1.12.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.1	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.8.0	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.13	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.13	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.21.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.21.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.

Classificações	Descrição
container-log4j	Altere os valores no arquivo container-log4j.properties YARN do Hadoop.
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4j-yarn-session.properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.

Classificações	Descrição
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.

Classificações	Descrição
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente do HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.

Classificações	Descrição
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.

Classificações	Descrição
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.

Classificações	Descrição
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.

Classificações	Descrição
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.20.1

Versões 5.20.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livvy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.20.1	emr-5.20.0	emr-5.19.1	emr-5.19.0
AWS SDK para Java	1.11.461	1.11.461	1.11.433	1.11.433
Python	2,7, 3,6	2,7, 3,6	2,7, 3,4	2,7, 3,4
Scala	2.11.12	2.11.12	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

	emr-5.20.1	emr-5.20.0	emr-5.19.1	emr-5.19.0
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	-	-	-	-
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

	emr-5.20.1	emr-5.20.0	emr-5.19.1	emr-5.19.0
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

Notas da versão 5.20.1

Esta é uma versão de patch para adicionar a autenticação AWS Signature versão 4 para solicitações ao Amazon S3. Todos os aplicativos e componentes são iguais aos da EMR versão anterior da Amazon.

Important

Nesta versão de lançamento, a Amazon EMR usa o AWS Signature versão 4 exclusivamente para autenticar solicitações para o Amazon S3. Para obter mais informações, consulte [Novidades](#).

Versões de componente 5.20.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR diferem das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
aws-sagemaker-spark-sdk	1.2.1	Amazon SageMaker Spark SDK
emr-ddb	4.7.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.5.1	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.10.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	1.2.0	EMRConector S3Select
emrfs	2.29.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.6.2	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.

Componente	Version (Versão)	Descrição
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.5-amzn-1	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.8.5-amzn-1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.8.5-amzn-1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-https-server	2.8.5-amzn-1	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.8.5-amzn-1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.8.5-amzn-1	YARNserviço para gerenciar contêineres em um nó individual.

Componente	Version (Versão)	Descrição
hadoop-yarn-resourcemanager	2.8.5-amzn-1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.5-amzn-1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.8	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.8	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.4.8	HBasecliente de linha de comando.
hbase-rest-server	1.4.8	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	1.4.8	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	2.3.4-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.4-amzn-0	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.

Componente	Version (Versão)	Descrição
hcatalog-webhcat-server	2.3.4-amzn-0	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.3.4-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.3.4-amzn-0	Cliente do Hive-hbase.
hive-metastore-server	2.3.4-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.4-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.3.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	0.9.4	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.5.0-incubating	REST interface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mahout-client	0.13.0	A biblioteca de Machine Learning.

Componente	Version (Versão)	Descrição
mxnet	1.3.1	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.2.88	Drivers NVIDIA e toolkit CUDA
oozie-client	5.0.0	O cliente da linha de comando do Oozie.
oozie-server	5.0.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	3.4.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.0- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.0- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.214	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.214	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.

Componente	Version (Versão)	Descrição
r	3.4.1	O projeto R para computação estatística
spark-client	2.4.0	Os clientes da linha de comando do Spark.
spark-history-server	2.4.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.4.0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.4.0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	1.12.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.1	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.8.0	O notebook baseado na web que permite a análise de dados interativa.

Componente	Version (Versão)	Descrição
zookeeper-server	3.4.13	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.13	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.20.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.20.1

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.

Classificações	Descrição
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.

Classificações	Descrição
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.

Classificações	Descrição
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente do HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.

Classificações	Descrição
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.

Classificações	Descrição
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.

Classificações	Descrição
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.

Classificações	Descrição
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.20.0

Versões 5.20.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)

- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.20.0	emr-5.19.1	emr-5.19.0	emr-5.18.1
AWS SDK para Java	1.11.461	1.11.433	1.11.433	1.11.393
Python	2,7, 3,6	2,7, 3,4	2,7, 3,4	2,7, 3,4
Scala	2.11.12	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	-	-	-	-

	emr-5.20.0	emr-5.19.1	emr-5.19.0	emr-5.18.1
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
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

Notas da versão 5.20.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.20.0 da Amazon. As alterações são referentes à versão 5.19.0.

Data da versão inicial: 18 de dezembro de 2018

Data da última atualização: 22 de janeiro de 2019

Atualizações

- 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

Novos atributos

- (22 de janeiro de 2019) O Kerberos na Amazon EMR foi aprimorado para oferecer suporte à autenticação de diretores externos. KDC Isso centraliza o gerenciamento principal porque vários clusters podem compartilhar um único externoKDC. Além disso, o externo KDC pode ter uma relação de confiança entre regiões com um domínio do Active Directory. Isso permite que todos os clusters autentiquem principais do Active Directory. Para obter mais informações, consulte [Usar a autenticação Kerberos](#) no Amazon EMR Management Guide.

Alterações, melhorias e problemas resolvidos

- Amazon Linux padrão AMI para Amazon EMR
 - O pacote Python3 foi atualizado do python 3.4 para 3.6.
- O committer EMRFS otimizado para S3
 - O committer EMRFS otimizado para S3 agora está habilitado por padrão, o que melhora o desempenho de gravação. Para ter mais informações, consulte [Use o committer EMRFS otimizado para S3](#).
- Hive
 - Portado para trás [HIVE-16686](#).
- Glue com Spark e Hive

- Na EMR versão 5.20.0 ou posterior, a remoção paralela de partições é ativada automaticamente para o Spark e o Hive quando o AWS Glue Data Catalog é usado como metastore. Essa alteração reduz significativamente o tempo de planejamento de consultas ao executar várias solicitações em paralelo para recuperar partições. O número total de segmentos que podem ser executados simultaneamente varia entre 1 e 10. O valor padrão é 5, que é uma configuração recomendada. Você pode alterá-lo especificando a propriedade `aws.glue.partition.num.segments` na classificação de configuração `hive-site`. Se ocorrer controle de utilização, você poderá desativar o atributo alterando o valor para 1. Para obter mais informações, consulte a [Estrutura de segmentos do AWS Glue](#).

Problemas conhecidos

- Hue (corrigido na EMR versão 5.24.0 da Amazon)
- O Hue executado na Amazon EMR não oferece suporte ao Solr. A partir da EMR versão 5.20.0 da Amazon, um problema de configuração incorreta faz com que o Solr seja ativado e uma mensagem de erro inofensiva apareça semelhante à seguinte:

```
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',))
```

Para evitar que a mensagem de erro do Solr seja exibida:

1. Conecte-se à linha de comando do nó primário usando SSH o.
2. Use um editor de texto para abrir o arquivo `hue.ini`. Por exemplo: .

```
sudo vim /etc/hue/conf/hue.ini
```

3. Pesquise o termo `appblacklist` e modifique a linha para o seguinte:

```
appblacklist = search
```

4. Salve as alterações e reinicie o Hue, conforme mostrado no exemplo a seguir:

```
sudo stop hue; sudo start hue
```

- Tez
 - Esse problema foi corrigido no Amazon EMR 5.22.0.

Quando você se conecta à interface do usuário do Tez em `http://MasterDNS:8080/tez-ui` por meio de uma SSH conexão com o nó primário do cluster, surge o erro “Falha na operação do adaptador - Timeline server () está fora de alcance. ATS “Está inativo ou não CORS está ativado” aparece ou as tarefas mostram inesperadamente N/A.

Isso é causado pela interface do usuário do Tez fazendo solicitações ao YARN Timeline Server usando, `localhost` em vez do nome do host, o nó primário. Como solução alternativa, um script está disponível para execução como ação ou etapa de bootstrap. O script atualiza o nome do host no arquivo `configs.env` do Tez. Para obter mais informações e a localização do script, consulte [Instruções de bootstrap](#).

- Nas EMR versões 5.19.0, 5.20.0 e 5.21.0 da Amazon, os rótulos dos YARN nós são armazenados em um diretório. HDFS Em algumas situações, isso leva a atrasos na inicialização do nó central causando, em seguida, tempo limite do cluster e falha na inicialização. A partir do Amazon EMR 5.22.0, esse problema foi resolvido. YARN os rótulos dos nós são armazenados no disco local de cada nó do cluster, evitando dependências em HDFS.
- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSH como usuário `hadoop` do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do `hadoop`.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo `keytab` está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

Versões de componente 5.20.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR diferem das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.2.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.7.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.5.1	Bibliotecas convenientes para o ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
emr-kinesis	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.10.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	1.2.0	EMRConector S3Select
emrfs	2.29.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.6.2	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.5-amzn-1	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.

Componente	Version (Versão)	Descrição
hadoop-hdfs-datanode	2.8.5-amzn-1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.8.5-amzn-1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.8.5-amzn-1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httfs-server	2.8.5-amzn-1	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.8.5-amzn-1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.8.5-amzn-1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.8.5-amzn-1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.5-amzn-1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.

Componente	Version (Versão)	Descrição
hbase-hmaster	1.4.8	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.8	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.4.8	HBase cliente de linha de comando.
hbase-rest-server	1.4.8	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.4.8	Serviço que fornece um endpoint Thrift para HBase.
hcatalog-client	2.3.4-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.4-amzn-0	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.4-amzn-0	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.3.4-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.3.4-amzn-0	Cliente do Hive-hbase.

Componente	Version (Versão)	Descrição
hive-metastore-server	2.3.4-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.4-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.3.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	0.9.4	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.5.0-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.3.1	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.2.88	Drivers NVIDIA e toolkit CUDA

Componente	Version (Versão)	Descrição
oozie-client	5.0.0	O cliente da linha de comando do Oozie.
oozie-server	5.0.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	3.4.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.0- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.0- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.214	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.214	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.1	O projeto R para computação estatística
spark-client	2.4.0	Os clientes da linha de comando do Spark.

Componente	Version (Versão)	Descrição
spark-history-server	2.4.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.4.0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.4.0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	1.12.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.9.1	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.
zeppelin-server	0.8.0	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.13	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.

Componente	Version (Versão)	Descrição
zookeeper-client	3.4.13	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.20.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.20.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Alterar as configurações de <code>log4j-cli.properties</code> no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.

Classificações	Descrição
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.

Classificações	Descrição
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.

Classificações	Descrição
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente do HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.

Classificações	Descrição
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.

Classificações	Descrição
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.

Classificações	Descrição
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.

Classificações	Descrição
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.19.1

Versões 5.19.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.19.1	emr-5.19.0	emr-5.18.1	emr-5.18.0
AWS SDK para Java	1.11.433	1.11.433	1.11.393	1.11.393

	emr-5.19.1	emr-5.19.0	emr-5.18.1	emr-5.18.0
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	1.6.1	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

	emr-5.19.1	emr-5.19.0	emr-5.18.1	emr-5.18.0
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	0.212	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

Notas da versão 5.19.1

Esta é uma versão de patch para adicionar a autenticação AWS Signature versão 4 para solicitações ao Amazon S3. Todos os aplicativos e componentes são iguais aos da EMR versão anterior da Amazon.

Important

Nesta versão de lançamento, a Amazon EMR usa o AWS Signature versão 4 exclusivamente para autenticar solicitações para o Amazon S3. Para obter mais informações, consulte [Novidades](#).

Versões de componente 5.19.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR

e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR diferem das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.2.0	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.7.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.5.1	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.10.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
<code>emr-s3-select</code>	1.1.0	EMRConector S3Select
<code>emrfs</code>	2.28.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
<code>flink-client</code>	1.6.1	Scripts do cliente da linha de comando e aplicativos do Apache Flink.

Componente	Version (Versão)	Descrição
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.5-amzn-0	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-0	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.8.5-amzn-0	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.8.5-amzn-0	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httpfs-server	2.8.5-amzn-0	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.8.5-amzn-0	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API

Componente	Version (Versão)	Descrição
hadoop-mapred	2.8.5-amzn-0	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.8.5-amzn-0	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.8.5-amzn-0	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.5-amzn-0	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.7	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.7	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.4.7	HBasecliente de linha de comando.
hbase-rest-server	1.4.7	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	1.4.7	Serviço que fornece um endpoint Thrift para. HBase

Componente	Version (Versão)	Descrição
hcatalog-client	2.3.3-amzn-2	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.3-amzn-2	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.3-amzn-2	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	2.3.3-amzn-2	O cliente da linha de comando do Hive.
hive-hbase	2.3.3-amzn-2	Cliente do Hive-hbase.
hive-metastore-server	2.3.3-amzn-2	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.3-amzn-2	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.2.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	0.9.4	Servidor de vários usuários para blocos de anotações Jupyter

Componente	Version (Versão)	Descrição
livy-server	0.5.0-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.3.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.2.88	Drivers NVIDIA e toolkit CUDA
oozie-client	5.0.0	O cliente da linha de comando do Oozie.
oozie-server	5.0.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	3.4.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.0- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.0- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API

Componente	Version (Versão)	Descrição
presto-coordinator	0.212	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.212	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.1	O projeto R para computação estatística
spark-client	2.3.2	Os clientes da linha de comando do Spark.
spark-history-server	2.3.2	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.3.2	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.3.2	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	1.11.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.

Componente	Version (Versão)	Descrição
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.8.0	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.13	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.13	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.19.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.19.1

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.

Classificações	Descrição
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.

Classificações	Descrição
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.

Classificações	Descrição
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
https-env	Mude os valores no HTTPFS ambiente.
https-site	Alterar os valores no arquivo https-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente do HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.

Classificações	Descrição
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.

Classificações	Descrição
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.

Classificações	Descrição
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.

Classificações	Descrição
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.19.0

Versões 5.19.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.19.0	emr-5.18.1	emr-5.18.0	emr-5.17.2
AWS SDK para Java	1.11.433	1.11.393	1.11.393	1.11.336
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
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

	emr-5.19.0	emr-5.18.1	emr-5.18.0	emr-5.17.2
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
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

Notas da versão 5.19.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.19.0 da Amazon. As alterações são referentes à versão 5.18.0.

Data da versão inicial: 7 de novembro de 2018

Data da última atualização: 19 de novembro de 2018

Atualizações

- 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 1.11.433

Novos atributos

- (19 de novembro de 2018) O EMR Notebooks é um ambiente gerenciado baseado no Jupyter Notebook. Ele suporta os kernels mágicos do Spark para PySpark, Spark, SQL Spark R e Scala. EMROs notebooks podem ser usados com clusters criados usando a EMR versão 5.18.0 e posterior da Amazon. Para obter mais informações, consulte [Usando EMR notebooks](#) no Guia de EMR gerenciamento da Amazon.
- O committer EMRFS otimizado para S3 está disponível ao gravar arquivos Parquet usando o Spark e. EMRFS Esse confirmador melhora a performance de gravação. Para ter mais informações, consulte [Use o committer EMRFS otimizado para S3](#).

Alterações, melhorias e problemas resolvidos

- YARN
 - Modificada a lógica que limita o processo mestre da aplicação à execução nos nós centrais. Essa funcionalidade agora usa o recurso de rótulos de YARN nós e as propriedades nas classificações `yarn-site` e `capacity-scheduler` de configuração. Para obter mais

informações, consulte <https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-plan-instances-guidelines.html#emr-plan-spot-YARN>.

- Amazon Linux padrão AMI para Amazon EMR
 - `ruby18`, `php56`, e `gcc48` não são mais instalados por padrão. Eles podem ser instalados, se desejado, usando `yum`.
 - A gem do ruby `aws-sdk` não é mais instalada por padrão. Ela pode ser instalada usando `gem install aws-sdk`, se desejado. Componentes específicos também podem ser instalados. Por exemplo, `gem install aws-sdk-s3`.

Problemas conhecidos

- EMRNotebooks — Em algumas circunstâncias, com vários editores de notebook abertos, o editor de notebook pode parecer incapaz de se conectar ao cluster. Se isso acontecer, limpe os cookies do navegador e reabra os editores de cadernos.
- CloudWatch ContainerPending Escalabilidade métrica e automática — (corrigida na versão 5.20.0) A Amazon EMR pode emitir um valor negativo para `ContainerPending` Se `ContainerPending` for usado em uma regra de escalabilidade automática, a escalabilidade automática não se comportará conforme esperado. Evite usar `ContainerPending` com escalabilidade automática.
- Nas EMR versões 5.19.0, 5.20.0 e 5.21.0 da Amazon, os rótulos dos YARN nós são armazenados em um diretório. HDFS Em algumas situações, isso leva a atrasos na inicialização do nó central causando, em seguida, tempo limite do cluster e falha na inicialização. A partir do Amazon EMR 5.22.0, esse problema foi resolvido. YARNos rótulos dos nós são armazenados no disco local de cada nó do cluster, evitando dependências emHDFS.

Versões de componente 5.19.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR diferem das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion`

começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado myapp-component com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como 2.2-amzn-2.

Componente	Version (Versão)	Descrição
aws-sagemaker-spark-sdk	1.2.0	Amazon SageMaker Spark SDK
emr-ddb	4.7.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.5.1	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.10.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	1.1.0	EMRConector S3Select
emrfs	2.28.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.6.1	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.

Componente	Version (Versão)	Descrição
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.5-amzn-0	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.5-amzn-0	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.8.5-amzn-0	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.8.5-amzn-0	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httpfs-server	2.8.5-amzn-0	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.8.5-amzn-0	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.8.5-amzn-0	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.

Componente	Version (Versão)	Descrição
hadoop-yarn-nodemanager	2.8.5-amzn-0	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.8.5-amzn-0	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.5-amzn-0	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.7	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.7	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.4.7	HBasecliente de linha de comando.
hbase-rest-server	1.4.7	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	1.4.7	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	2.3.3-amzn-2	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.

Componente	Version (Versão)	Descrição
hcatalog-server	2.3.3-amzn-2	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.3-amzn-2	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	2.3.3-amzn-2	O cliente da linha de comando do Hive.
hive-hbase	2.3.3-amzn-2	Cliente do Hive-hbase.
hive-metastore-server	2.3.3-amzn-2	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.3-amzn-2	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.2.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	0.9.4	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.5.0-incubating	RESTinterface para interagir com o Apache Spark

Componente	Version (Versão)	Descrição
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.3.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.2.88	Drivers NVIDIA e toolkit CUDA
oozie-client	5.0.0	O cliente da linha de comando do Oozie.
oozie-server	5.0.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	3.4.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.0- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.0- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.212	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.

Componente	Version (Versão)	Descrição
presto-worker	0.212	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.1	O projeto R para computação estatística
spark-client	2.3.2	Os clientes da linha de comando do Spark.
spark-history-server	2.3.2	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.3.2	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.3.2	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	1.11.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.

Componente	Version (Versão)	Descrição
zeppelin-server	0.8.0	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.13	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.13	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.19.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.19.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .

Classificações	Descrição
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.

Classificações	Descrição
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.

Classificações	Descrição
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente do HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.

Classificações	Descrição
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.

Classificações	Descrição
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-memory	Alterar os valores no arquivo memory.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.

Classificações	Descrição
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
presto-connector-tpcds	Alterar os valores no arquivo tpcds.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.

Classificações	Descrição
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.18.1

Versões 5.18.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.18.1	emr-5.18.0	emr-5.17.2	emr-5.17.1
AWS SDK para Java	1.11.393	1.11.393	1.11.336	1.11.336
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	1.6.0	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	-	-	-	-

	emr-5.18.1	emr-5.18.0	emr-5.17.2	emr-5.17.1
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	0.210	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

Notas da versão 5.18.1

Esta é uma versão de patch para adicionar a autenticação AWS Signature versão 4 para solicitações ao Amazon S3. Todos os aplicativos e componentes são iguais aos da EMR versão anterior da Amazon.

⚠ Important

Nesta versão de lançamento, a Amazon EMR usa o AWS Signature versão 4 exclusivamente para autenticar solicitações para o Amazon S3. Para obter mais informações, consulte [Novidades](#).

Versões de componente 5.18.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR diferem das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.1.3	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.6.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.5.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
emr-s3-dist-cp	2.10.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	1.1.0	EMRConector S3Select
emrfs	2.27.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.6.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.4-amzn-1	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.4-amzn-1	HDFSserviço em nível de nó para armazenar blocos.

Componente	Version (Versão)	Descrição
hadoop-hdfs-library	2.8.4-amzn-1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.8.4-amzn-1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httfs-server	2.8.4-amzn-1	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.8.4-amzn-1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.8.4-amzn-1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.8.4-amzn-1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.8.4-amzn-1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.4-amzn-1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.7	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.

Componente	Version (Versão)	Descrição
hbase-region-server	1.4.7	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.4.7	HBase cliente de linha de comando.
hbase-rest-server	1.4.7	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.4.7	Serviço que fornece um endpoint Thrift para HBase
hcatalog-client	2.3.3-amzn-2	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.3-amzn-2	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.3-amzn-2	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.3.3-amzn-2	O cliente da linha de comando do Hive.
hive-hbase	2.3.3-amzn-2	Cliente do Hive-hbase.
hive-metastore-server	2.3.3-amzn-2	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL

Componente	Version (Versão)	Descrição
hive-server2	2.3.3-amzn-2	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.2.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	0.8.1	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.5.0-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.2.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.2.88	Drivers NVIDIA e toolkit CUDA
oozie-client	5.0.0	O cliente da linha de comando do Oozie.
oozie-server	5.0.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.

Componente	Version (Versão)	Descrição
opencv	3.4.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.0- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.0- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.210	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.210	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.1	O projeto R para computação estatística
spark-client	2.3.2	Os clientes da linha de comando do Spark.
spark-history-server	2.3.2	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.3.2	Mecanismo de execução na memória para YARN.

Componente	Version (Versão)	Descrição
spark-yarn-slave	2.3.2	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	1.9.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.8.0	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.12	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.12	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.18.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.18.1

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.
container-log4j	Altere os valores no arquivo container-log4j.properties YARN do Hadoop.
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4j-yarn-session.properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase

Classificações	Descrição
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.

Classificações	Descrição
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente do HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.

Classificações	Descrição
jupyter-notebook-conf	Alterar os valores no arquivo <code>jupyter_notebook_config.py</code> do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo <code>jupyterhub_config.py</code> do.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.
jupyter-sparkmagic-conf	Altere os valores no arquivo <code>config.json</code> do Sparkmagic.
livy-conf	Alterar os valores no arquivo <code>livy.conf</code> do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de <code>log4j.properties</code> no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo <code>mapred-site.xml</code> do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo <code>oozie-log4j.properties</code> do Oozie.
oozie-site	Alterar os valores no arquivo <code>oozie-site.xml</code> do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo <code>hadoop-metrics2-hbase.properties</code> do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo <code>hbase-site.xml</code> do Phoenix.

Classificações	Descrição
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.

Classificações	Descrição
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.

Classificações	Descrição
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oracoop-site	Altere os valores no arquivo oracoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.18.0

Versões 5.18.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.18.0	emr-5.17.2	emr-5.17.1	emr-5.17.0
AWS SDK para Java	1.11.393	1.11.336	1.11.336	1.11.336
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

Notas da versão 5.18.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.18.0 da Amazon. As alterações são referentes à versão 5.17.0.

Data da versão inicial: 24 de outubro de 2018

Atualizações

- Flink 1.6.0
- HBase1.4.7
- Presto 0.210
- Spark 2.3.2
- Zeppelin 0.8.0

Novos atributos

- A partir do Amazon EMR 5.18.0, você pode usar o repositório de EMR artefatos da Amazon para criar seu código de trabalho com base nas versões exatas das bibliotecas e dependências que estão disponíveis em versões específicas da Amazon. Para ter mais informações, consulte [Verificar dependências usando o repositório de artefatos do Amazon EMR](#).

Alterações, melhorias e problemas resolvidos

- Hive
 - Adicionado suporte para o S3 Select. Para ter mais informações, consulte [Usar o S3 Select com o Hive para melhorar a performance](#).
- Presto

- Adicionado suporte para o [S3 Select](#) Pushdown. Para ter mais informações, consulte [Usar S3 Select Pushdown com o Presto para melhorar a performance](#).
- Spark
 - A configuração log4j padrão do Spark foi alterada para lançar logs de contêineres por hora para trabalhos de streaming do Spark. Isso ajuda a evitar a exclusão de logs de trabalhos de streaming do Spark de execução prolongada.

Versões de componente 5.18.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR diferem das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.1.3	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.6.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.5.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
emr-s3-dist-cp	2.10.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	1.1.0	EMRConector S3Select
emrfs	2.27.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.6.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.4-amzn-1	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.4-amzn-1	HDFSserviço em nível de nó para armazenar blocos.

Componente	Version (Versão)	Descrição
hadoop-hdfs-library	2.8.4-amzn-1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.8.4-amzn-1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httfs-server	2.8.4-amzn-1	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.8.4-amzn-1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.8.4-amzn-1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.8.4-amzn-1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.8.4-amzn-1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.4-amzn-1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.7	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.

Componente	Version (Versão)	Descrição
hbase-region-server	1.4.7	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.4.7	HBase cliente de linha de comando.
hbase-rest-server	1.4.7	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.4.7	Serviço que fornece um endpoint Thrift para HBase
hcatalog-client	2.3.3-amzn-2	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.3-amzn-2	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.3-amzn-2	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.3.3-amzn-2	O cliente da linha de comando do Hive.
hive-hbase	2.3.3-amzn-2	Cliente do Hive-hbase.
hive-metastore-server	2.3.3-amzn-2	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL

Componente	Version (Versão)	Descrição
hive-server2	2.3.3-amzn-2	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.2.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	0.8.1	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.5.0-incubating	RESTinterface para interagir com o Apache Spark
nginx	1.12.1	nginx [engine x] é HTTP um servidor proxy reverso
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.2.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.2.88	Drivers NVIDIA e toolkit CUDA
oozie-client	5.0.0	O cliente da linha de comando do Oozie.
oozie-server	5.0.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.

Componente	Version (Versão)	Descrição
opencv	3.4.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.0- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.0- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.210	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.210	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.1	O projeto R para computação estatística
spark-client	2.3.2	Os clientes da linha de comando do Spark.
spark-history-server	2.3.2	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.3.2	Mecanismo de execução na memória para YARN.

Componente	Version (Versão)	Descrição
spark-yarn-slave	2.3.2	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	1.9.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.8.0	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.12	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.12	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.18.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.18.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.
container-log4j	Altere os valores no arquivo container-log4j.properties YARN do Hadoop.
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4j-yarn-session.properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase

Classificações	Descrição
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.

Classificações	Descrição
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente do HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.

Classificações	Descrição
jupyter-notebook-conf	Alterar os valores no arquivo <code>jupyter_notebook_config.py</code> do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo <code>jupyterhub_config.py</code> do.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.
jupyter-sparkmagic-conf	Altere os valores no arquivo <code>config.json</code> do Sparkmagic.
livy-conf	Alterar os valores no arquivo <code>livy.conf</code> do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de <code>log4j.properties</code> no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo <code>mapred-site.xml</code> do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo <code>oozie-log4j.properties</code> do Oozie.
oozie-site	Alterar os valores no arquivo <code>oozie-site.xml</code> do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo <code>hadoop-metrics2-hbase.properties</code> do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo <code>hbase-site.xml</code> do Phoenix.

Classificações	Descrição
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.

Classificações	Descrição
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.

Classificações	Descrição
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oracoop-site	Altere os valores no arquivo oracoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.17.2

Versões 5.17.2 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.17.2	emr-5.17.1	emr-5.17.0	emr-5.16.1
AWS SDK para Java	1.11.336	1.11.336	1.11.336	1.11.336
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

	emr-5.17.2	emr-5.17.1	emr-5.17.0	emr-5.16.1
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	0.206	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

	emr-5.17.2	emr-5.17.1	emr-5.17.0	emr-5.16.1
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

Notas da versão 5.17.2

Esta é uma versão de patch para adicionar a autenticação AWS Signature versão 4 para solicitações ao Amazon S3. Todos os aplicativos e componentes são iguais aos da EMR versão anterior da Amazon.

Important

Nesta versão de lançamento, a Amazon EMR usa o AWS Signature versão 4 exclusivamente para autenticar solicitações para o Amazon S3. Para obter mais informações, consulte [Novidades](#).

Versões de componente 5.17.2

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR diferem das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
aws-sagemaker-spark-sdk	1.1.3	Amazon SageMaker Spark SDK
emr-ddb	4.6.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.5.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.10.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emr-s3-select	1.0.0	EMRConector S3Select
emrfs	2.26.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.5.2	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.

Componente	Version (Versão)	Descrição
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.4-amzn-1	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.4-amzn-1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.8.4-amzn-1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.8.4-amzn-1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-https-server	2.8.4-amzn-1	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.8.4-amzn-1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.8.4-amzn-1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.8.4-amzn-1	YARNserviço para gerenciar contêineres em um nó individual.

Componente	Version (Versão)	Descrição
hadoop-yarn-resourcemanager	2.8.4-amzn-1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.4-amzn-1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.6	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.6	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.4.6	HBasecliente de linha de comando.
hbase-rest-server	1.4.6	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	1.4.6	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	2.3.3-amzn-1	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.3-amzn-1	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.

Componente	Version (Versão)	Descrição
hcatalog-webhcat-server	2.3.3-amzn-1	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.3.3-amzn-1	O cliente da linha de comando do Hive.
hive-hbase	2.3.3-amzn-1	Cliente do Hive-hbase.
hive-metastore-server	2.3.3-amzn-1	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.3-amzn-1	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.2.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	0.8.1	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.5.0-incubating	REST interface para interagir com o Apache Spark
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.2.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.

Componente	Version (Versão)	Descrição
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.2.88	Drivers NVIDIA e toolkit CUDA
oozie-client	5.0.0	O cliente da linha de comando do Oozie.
oozie-server	5.0.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	3.4.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.0- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.0- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.206	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.206	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.1	O projeto R para computação estatística

Componente	Version (Versão)	Descrição
spark-client	2.3.1	Os clientes da linha de comando do Spark.
spark-history-server	2.3.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.3.1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.3.1	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	1.9.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.
zeppelin-server	0.7.3	O notebook baseado na web que permite a análise de dados interativa.

Componente	Version (Versão)	Descrição
zookeeper-server	3.4.12	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.12	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.17.2

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.17.2

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.

Classificações	Descrição
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.

Classificações	Descrição
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.

Classificações	Descrição
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente do HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.

Classificações	Descrição
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.

Classificações	Descrição
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.

Classificações	Descrição
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.

Classificações	Descrição
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.17.1

Versões 5.17.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.17.1	emr-5.17.0	emr-5.16.1	emr-5.16.0
AWS SDK para Java	1.11.336	1.11.336	1.11.336	1.11.336
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
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-5.17.1	emr-5.17.0	emr-5.16.1	emr-5.16.0
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
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

Notas da versão 5.17.1

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.17.1 da Amazon. As alterações são referentes à versão 5.17.0.

Data da versão inicial: 18 de julho de 2019

Alterações, melhorias e problemas resolvidos

- Atualizou o Amazon Linux padrão AMI para Amazon EMR para incluir atualizações importantes de segurança do kernel Linux, incluindo o problema de TCP SACK negação de serviço ([AWS-2019-005](#)).

Versões de componente 5.17.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR diferem das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.1.3	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.6.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.5.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.10.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
<code>emr-s3-select</code>	1.0.0	EMRConector S3Select
<code>emrfs</code>	2.26.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
flink-client	1.5.2	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.4-amzn-1	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.4-amzn-1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.8.4-amzn-1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.8.4-amzn-1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-https-server	2.8.4-amzn-1	HTTPendpoint para HDFS operações.

Componente	Version (Versão)	Descrição
hadoop-kms-server	2.8.4-amzn-1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.8.4-amzn-1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.8.4-amzn-1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.8.4-amzn-1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.4-amzn-1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.6	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.6	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.4.6	HBasecliente de linha de comando.
hbase-rest-server	1.4.6	Serviço que fornece um RESTful HTTP endpoint paraHBase.

Componente	Version (Versão)	Descrição
hbase-thrift-server	1.4.6	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	2.3.3-amzn-1	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.3-amzn-1	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.3-amzn-1	HTTPendpoint fornecendo o uma REST interface paraHCatalog.
hive-client	2.3.3-amzn-1	O cliente da linha de comando do Hive.
hive-hbase	2.3.3-amzn-1	Cliente do Hive-hbase.
hive-metastore-server	2.3.3-amzn-1	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.3-amzn-1	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.2.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop

Componente	Version (Versão)	Descrição
jupyterhub	0.8.1	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.5.0-incubating	RESTinterface para interagir com o Apache Spark
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.2.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.2.88	Drivers NVIDIA e toolkit CUDA
oozie-client	5.0.0	O cliente da linha de comando do Oozie.
oozie-server	5.0.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	3.4.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.0- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.0- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API

Componente	Version (Versão)	Descrição
presto-coordinator	0.206	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.206	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.1	O projeto R para computação estatística
spark-client	2.3.1	Os clientes da linha de comando do Spark.
spark-history-server	2.3.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.3.1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.3.1	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	1.9.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.

Componente	Version (Versão)	Descrição
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.7.3	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.12	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.12	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.17.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.17.1

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.

Classificações	Descrição
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.

Classificações	Descrição
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.

Classificações	Descrição
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
https-env	Mude os valores no HTTPFS ambiente.
https-site	Alterar os valores no arquivo https-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente do HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.

Classificações	Descrição
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.

Classificações	Descrição
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.

Classificações	Descrição
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.

Classificações	Descrição
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.17.0

Versões 5.17.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [TensorFlow](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)

- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.17.0	emr-5.16.1	emr-5.16.0	emr-5.15.1
AWS SDK para Java	1.11.336	1.11.336	1.11.336	1.11.333
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	-	-	-	-

	emr-5.17.0	emr-5.16.1	emr-5.16.0	emr-5.15.1
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
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

Notas da versão 5.17.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.17.0 da Amazon. As alterações são referentes à versão 5.16.0.

Data da versão inicial: 30 de agosto de 2018

Atualizações

- Flink 1.5.2
- HBase1.4.6
- Presto 0.206

Novos atributos

- Adicionado suporte para Tensorflow. Para ter mais informações, consulte [TensorFlow](#).

Alterações, melhorias e problemas resolvidos

- JupyterHub
 - Adicionado suporte para a persistência de cadernos no Amazon S3. Para ter mais informações, consulte [Configurar a persistência de cadernos no Amazon S3](#).
- Spark
 - Adicionado suporte para o [S3 Select](#). Para ter mais informações, consulte [Usar o S3 Select com Spark para melhorar a performance das consultas](#).
- Resolveu os problemas com as métricas do Cloudwatch e o recurso de escalabilidade automática na Amazon EMR versão 5.14.0, 5.15.0 ou 5.16.0.

Problemas conhecidos

- Quando você cria um cluster kerberizado com o Livy instalado, o Livy apresenta falha com um erro em que a autenticação simples não está habilitada. A reinicialização do servidor do Livy resolve o problema. Como solução alternativa, adicione uma etapa durante a criação do cluster que execute `sudo restart livy-server` no nó primário.
- Se você usa um Amazon Linux personalizado AMI baseado em um Amazon Linux AMI com uma data de criação de 11/08/2018, o servidor Oozie falhará ao iniciar. Se você usa o Oozie, crie um personalizado AMI com base em um Amazon Linux AMI ID com uma data de criação diferente. Você pode usar o AWS CLI comando a seguir para retornar uma lista de imagens IDs para todo o HVM Amazon Linux AMIs com uma versão 2018.03, junto com a data de lançamento, para que você possa escolher um Amazon Linux apropriado AMI como sua base. MyRegion Substitua pelo seu identificador de região, como 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
```

Versões de componente 5.17.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR diferem das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.1.3	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.6.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.5.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.10.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.

Componente	Version (Versão)	Descrição
emr-s3-select	1.0.0	EMRConector S3Select
emrfs	2.26.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.5.2	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.4-amzn-1	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.4-amzn-1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.8.4-amzn-1	HDFScliente e biblioteca de linha de comando

Componente	Version (Versão)	Descrição
hadoop-hdfs-namenode	2.8.4-amzn-1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httfs-server	2.8.4-amzn-1	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.8.4-amzn-1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.8.4-amzn-1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.8.4-amzn-1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.8.4-amzn-1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.4-amzn-1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.6	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.6	Serviço para atender a uma ou mais HBase regiões.

Componente	Version (Versão)	Descrição
hbase-client	1.4.6	HBase cliente de linha de comando.
hbase-rest-server	1.4.6	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.4.6	Serviço que fornece um endpoint Thrift para HBase
hcatalog-client	2.3.3-amzn-1	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.3-amzn-1	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.3-amzn-1	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.3.3-amzn-1	O cliente da linha de comando do Hive.
hive-hbase	2.3.3-amzn-1	Cliente do Hive-hbase.
hive-metastore-server	2.3.3-amzn-1	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL

Componente	Version (Versão)	Descrição
hive-server2	2.3.3-amzn-1	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.2.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	0.8.1	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.5.0-incubating	RESTinterface para interagir com o Apache Spark
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.2.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.2.88	Drivers NVIDIA e toolkit CUDA
oozie-client	5.0.0	O cliente da linha de comando do Oozie.
oozie-server	5.0.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	3.4.0	Biblioteca de Visão Computacional de Código Aberto.

Componente	Version (Versão)	Descrição
phoenix-library	4.14.0- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.0- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.206	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.206	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.1	O projeto R para computação estatística
spark-client	2.3.1	Os clientes da linha de comando do Spark.
spark-history-server	2.3.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.3.1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.3.1	Bibliotecas do Apache Spark necessárias para escravos. YARN

Componente	Version (Versão)	Descrição
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tensorflow	1.9.0	TensorFlow biblioteca de software de código aberto para computação numérica de alto desempenho.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.7.3	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.12	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.12	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.17.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.17.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.
container-log4j	Altere os valores no arquivo container-log4j.properties YARN do Hadoop.
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4j-yarn-session.properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase

Classificações	Descrição
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.

Classificações	Descrição
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente do HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.

Classificações	Descrição
jupyter-notebook-conf	Alterar os valores no arquivo <code>jupyter_notebook_config.py</code> do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo <code>jupyterhub_config.py</code> do.
jupyter-s3-conf	Configurar a persistência do notebook Jupyter S3.
jupyter-sparkmagic-conf	Altere os valores no arquivo <code>config.json</code> do Sparkmagic.
livy-conf	Alterar os valores no arquivo <code>livy.conf</code> do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de <code>log4j.properties</code> no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo <code>mapred-site.xml</code> do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo <code>oozie-log4j.properties</code> do Oozie.
oozie-site	Alterar os valores no arquivo <code>oozie-site.xml</code> do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo <code>hadoop-metrics2-hbase.properties</code> do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo <code>hbase-site.xml</code> do Phoenix.

Classificações	Descrição
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.

Classificações	Descrição
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.

Classificações	Descrição
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oracoop-site	Altere os valores no arquivo oracoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.16.1

Versões 5.16.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livvy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.16.1	emr-5.16.0	emr-5.15.1	emr-5.15.0
AWS SDK para Java	1.11.336	1.11.336	1.11.333	1.11.333
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

Notas da versão 5.16.1

Esta é uma versão de patch para adicionar a autenticação AWS Signature versão 4 para solicitações ao Amazon S3. Todos os aplicativos e componentes são iguais aos da EMR versão anterior da Amazon.

Important

Nesta versão de lançamento, a Amazon EMR usa o AWS Signature versão 4 exclusivamente para autenticar solicitações para o Amazon S3. Para obter mais informações, consulte [Novidades](#).

Versões de componente 5.16.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR diferem das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
aws-sagemaker-spark-sdk	1.1.0	Amazon SageMaker Spark SDK
emr-ddb	4.6.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.4.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.10.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.25.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.5.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.

Componente	Version (Versão)	Descrição
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.4-amzn-0	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.4-amzn-0	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.8.4-amzn-0	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.8.4-amzn-0	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-https-server	2.8.4-amzn-0	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.8.4-amzn-0	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.8.4-amzn-0	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.8.4-amzn-0	YARNserviço para gerenciar contêineres em um nó individual.

Componente	Version (Versão)	Descrição
hadoop-yarn-resourcemanager	2.8.4-amzn-0	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.4-amzn-0	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.4	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.4	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.4.4	HBasecliente de linha de comando.
hbase-rest-server	1.4.4	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	1.4.4	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	2.3.3-amzn-1	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.3-amzn-1	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.

Componente	Version (Versão)	Descrição
hcatalog-webhcat-server	2.3.3-amzn-1	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.3.3-amzn-1	O cliente da linha de comando do Hive.
hive-hbase	2.3.3-amzn-1	Cliente do Hive-hbase.
hive-metastore-server	2.3.3-amzn-1	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.3-amzn-1	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.2.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	0.8.1	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.5.0-incubating	REST interface para interagir com o Apache Spark
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.2.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.

Componente	Version (Versão)	Descrição
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.2.88	Drivers NVIDIA e toolkit CUDA
oozie-client	5.0.0	O cliente da linha de comando do Oozie.
oozie-server	5.0.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	3.4.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.0- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.0- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.203	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.203	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.1	O projeto R para computação estatística

Componente	Version (Versão)	Descrição
spark-client	2.3.1	Os clientes da linha de comando do Spark.
spark-history-server	2.3.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.3.1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.3.1	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.
zeppelin-server	0.7.3	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.12	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.12	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.16.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.16.1

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Alterar as configurações de <code>log4j-cli.properties</code> no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop

Classificações	Descrição
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure as zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.

Classificações	Descrição
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente do HadoopKMS.

Classificações	Descrição
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.

Classificações	Descrição
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.

Classificações	Descrição
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.

Classificações	Descrição
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.16.0

Versões 5.16.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.16.0	emr-5.15.1	emr-5.15.0	emr-5.14.2
AWS SDK para Java	1.11.336	1.11.333	1.11.333	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

	emr-5.16.0	emr-5.15.1	emr-5.15.0	emr-5.14.2
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
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

	emr-5.16.0	emr-5.15.1	emr-5.15.0	emr-5.14.2
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

Notas da versão 5.16.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.16.0 da Amazon. As alterações são referentes à versão 5.15.0.

Data da versão inicial: 19 de julho de 2018

Atualizações

- 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 1.11.336
- CUDA9.2
- Controlador Redshift 1.2.15.1025 JDBC

Alterações, melhorias e problemas resolvidos

- HBase
 - [Porta traseira -20723 HBASE](#)
- Presto
 - Alterações na configuração para oferecer suporte à LDAP autenticação. Para ter mais informações, consulte [Usando a LDAP autenticação para o Presto na Amazon EMR](#).

- Spark
 - [A versão 2.3.1 do Apache Spark, disponível a partir da EMR versão 5.16.0 da Amazon, endereça -2018-8024 e -2018-1334. CVE CVE](#) Recomendamos que você migre as versões anteriores do Spark para a versão 2.3.1 ou posteriores.

Problemas conhecidos

- Essa versão não é compatível com os tipos de instância c1.medium ou m1.small. Os clusters que usam qualquer um desses tipos de instância não são iniciados. Como solução alternativa, especifique um tipo de instância diferente ou use uma versão diferente.
- Quando você cria um cluster kerberizado com o Livy instalado, o Livy apresenta falha com um erro em que a autenticação simples não está habilitada. A reinicialização do servidor do Livy resolve o problema. Como solução alternativa, adicione uma etapa durante a criação do cluster que execute `sudo restart livy-server` no nó primário.
- Depois que o nó primário for reinicializado ou o controlador de instância for reiniciado, as CloudWatch métricas não serão coletadas e o recurso de escalabilidade automática não estará disponível nas EMR versões 5.14.0, 5.15.0 ou 5.16.0 da Amazon. Esse problema foi corrigido no Amazon EMR 5.17.0.

Versões de componente 5.16.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR diferem das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
aws-sagemaker-spark-sdk	1.1.0	Amazon SageMaker Spark SDK
emr-ddb	4.6.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.4.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.10.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.25.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.5.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.

Componente	Version (Versão)	Descrição
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.4-amzn-0	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.4-amzn-0	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.8.4-amzn-0	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.8.4-amzn-0	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-https-server	2.8.4-amzn-0	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.8.4-amzn-0	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.8.4-amzn-0	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.8.4-amzn-0	YARNserviço para gerenciar contêineres em um nó individual.

Componente	Version (Versão)	Descrição
hadoop-yarn-resourcemanager	2.8.4-amzn-0	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.4-amzn-0	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.4	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.4	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.4.4	HBasecliente de linha de comando.
hbase-rest-server	1.4.4	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	1.4.4	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	2.3.3-amzn-1	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.3-amzn-1	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.

Componente	Version (Versão)	Descrição
hcatalog-webhcat-server	2.3.3-amzn-1	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.3.3-amzn-1	O cliente da linha de comando do Hive.
hive-hbase	2.3.3-amzn-1	Cliente do Hive-hbase.
hive-metastore-server	2.3.3-amzn-1	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.3-amzn-1	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.2.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	0.8.1	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.5.0-incubating	REST interface para interagir com o Apache Spark
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.2.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.

Componente	Version (Versão)	Descrição
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.2.88	Drivers NVIDIA e toolkit CUDA
oozie-client	5.0.0	O cliente da linha de comando do Oozie.
oozie-server	5.0.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	3.4.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.14.0- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.14.0- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.203	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.203	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.1	O projeto R para computação estatística

Componente	Version (Versão)	Descrição
spark-client	2.3.1	Os clientes da linha de comando do Spark.
spark-history-server	2.3.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.3.1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.3.1	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.
zeppelin-server	0.7.3	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.12	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.12	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.16.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.16.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Alterar as configurações de <code>log4j-cli.properties</code> no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop

Classificações	Descrição
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure as zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.

Classificações	Descrição
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente do HadoopKMS.

Classificações	Descrição
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.

Classificações	Descrição
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-password-authenticator	Altere os valores no arquivo password-authenticator.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.

Classificações	Descrição
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.

Classificações	Descrição
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.15.1

Versões 5.15.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.15.1	emr-5.15.0	emr-5.14.2	emr-5.14.1
AWS SDK para Java	1.11.333	1.11.333	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

	emr-5.15.1	emr-5.15.0	emr-5.14.2	emr-5.14.1
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
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

	emr-5.15.1	emr-5.15.0	emr-5.14.2	emr-5.14.1
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

Notas da versão 5.15.1

Esta é uma versão de patch para adicionar a autenticação AWS Signature versão 4 para solicitações ao Amazon S3. Todos os aplicativos e componentes são iguais aos da EMR versão anterior da Amazon.

Important

Nesta versão de lançamento, a Amazon EMR usa o AWS Signature versão 4 exclusivamente para autenticar solicitações para o Amazon S3. Para obter mais informações, consulte [Novidades](#).

Versões de componente 5.15.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR diferem das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
aws-sagemaker-spark-sdk	1.0.1	Amazon SageMaker Spark SDK
emr-ddb	4.5.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.4.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.10.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.24.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.4.2	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.

Componente	Version (Versão)	Descrição
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.3-amzn-1	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.3-amzn-1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.8.3-amzn-1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.8.3-amzn-1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-https-server	2.8.3-amzn-1	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.8.3-amzn-1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.8.3-amzn-1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.8.3-amzn-1	YARNserviço para gerenciar contêineres em um nó individual.

Componente	Version (Versão)	Descrição
hadoop-yarn-resourcemanager	2.8.3-amzn-1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.3-amzn-1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.4	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.4	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.4.4	HBasecliente de linha de comando.
hbase-rest-server	1.4.4	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	1.4.4	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	2.3.3-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.3-amzn-0	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.

Componente	Version (Versão)	Descrição
hcatalog-webhcat-server	2.3.3-amzn-0	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.3.3-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.3.3-amzn-0	Cliente do Hive-hbase.
hive-metastore-server	2.3.3-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.3-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.2.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	0.8.1	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.4.0-incubating	REST interface para interagir com o Apache Spark
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.1.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.

Componente	Version (Versão)	Descrição
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.1.85	Drivers NVIDIA e toolkit CUDA
oozie-client	5.0.0	O cliente da linha de comando do Oozie.
oozie-server	5.0.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	3.4.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.13.0- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.13.0- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.194	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.194	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.1	O projeto R para computação estatística

Componente	Version (Versão)	Descrição
spark-client	2.3.0	Os clientes da linha de comando do Spark.
spark-history-server	2.3.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.3.0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.3.0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.
zeppelin-server	0.7.3	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.12	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.12	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.15.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.15.1

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Alterar as configurações de <code>log4j-cli.properties</code> no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop

Classificações	Descrição
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure as zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.

Classificações	Descrição
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente do HadoopKMS.

Classificações	Descrição
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.

Classificações	Descrição
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.

Classificações	Descrição
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.

Classificações	Descrição
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oracoop-site	Altere os valores no arquivo oracoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.15.0

Versões 5.15.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.15.0	emr-5.14.2	emr-5.14.1	emr-5.14.0
AWS SDK para Java	1.11.333	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

	emr-5.15.0	emr-5.14.2	emr-5.14.1	emr-5.14.0
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	-	-	-	-
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

	emr-5.15.0	emr-5.14.2	emr-5.14.1	emr-5.14.0
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

Notas da versão 5.15.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.15.0 da Amazon. As alterações são referentes à versão 5.14.0.

Data da versão inicial: 21 de junho de 2018

Atualizações

- Atualizado HBase para 1.4.4
- Atualizado Hive para 2.3.3
- Atualizado Hue para 4.2.0
- Atualizado Oozie para 5.0.0
- Atualizado Zookeeper para 3.4.12
- Atualizado para AWS SDK 1.11.333

Alterações, melhorias e problemas resolvidos

- Hive
 - [Portado para trás - 18069 HIVE](#)
- Hue
 - Atualizado o Hue para se autenticar corretamente com o Livy quando o Kerberos está habilitado. Agora, o Livy é suportado ao usar o Kerberos com a Amazon. EMR
- JupyterHub
 - Atualizado JupyterHub para que a Amazon EMR instale bibliotecas de LDAP clientes por padrão.
 - Corrigido um erro no script que gera certificados autoassinados.

Problemas conhecidos

- Essa versão não é compatível com os tipos de instância c1.medium ou m1.small. Os clusters que usam qualquer um desses tipos de instância não são iniciados. Como solução alternativa, especifique um tipo de instância diferente ou use uma versão diferente.
- Depois que o nó primário for reinicializado ou o controlador de instância for reiniciado, as CloudWatch métricas não serão coletadas e o recurso de escalabilidade automática não estará disponível nas EMR versões 5.14.0, 5.15.0 ou 5.16.0 da Amazon. Esse problema foi corrigido no Amazon EMR 5.17.0.

Versões de componente 5.15.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR diferem das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.0.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.5.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.4.0	Bibliotecas convenientes para o ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
emr-kinesis	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.10.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.24.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.4.2	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.3-amzn-1	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.3-amzn-1	HDFSserviço em nível de nó para armazenar blocos.

Componente	Version (Versão)	Descrição
hadoop-hdfs-library	2.8.3-amzn-1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.8.3-amzn-1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httfs-server	2.8.3-amzn-1	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.8.3-amzn-1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.8.3-amzn-1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.8.3-amzn-1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.8.3-amzn-1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.3-amzn-1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.4	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.

Componente	Version (Versão)	Descrição
hbase-region-server	1.4.4	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.4.4	HBase cliente de linha de comando.
hbase-rest-server	1.4.4	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.4.4	Serviço que fornece um endpoint Thrift para HBase.
hcatalog-client	2.3.3-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.3-amzn-0	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.3-amzn-0	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.3.3-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.3.3-amzn-0	Cliente do Hive-hbase.
hive-metastore-server	2.3.3-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL

Componente	Version (Versão)	Descrição
hive-server2	2.3.3-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.2.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	0.8.1	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.4.0-incubating	RESTinterface para interagir com o Apache Spark
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.1.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.1.85	Drivers NVIDIA e toolkit CUDA
oozie-client	5.0.0	O cliente da linha de comando do Oozie.
oozie-server	5.0.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	3.4.0	Biblioteca de Visão Computacional de Código Aberto.

Componente	Version (Versão)	Descrição
phoenix-library	4.13.0- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.13.0- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.194	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.194	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.1	O projeto R para computação estatística
spark-client	2.3.0	Os clientes da linha de comando do Spark.
spark-history-server	2.3.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.3.0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.3.0	Bibliotecas do Apache Spark necessárias para escravos. YARN

Componente	Version (Versão)	Descrição
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.7.3	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.12	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.12	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.15.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.15.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.

Classificações	Descrição
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.

Classificações	Descrição
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure as zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.

Classificações	Descrição
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente do HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.

Classificações	Descrição
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.

Classificações	Descrição
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.

Classificações	Descrição
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.

Classificações	Descrição
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.14.2

Versões 5.14.2 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.14.2	emr-5.14.1	emr-5.14.0	emr-5.13.1
AWS SDK para 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

Notas da versão 5.14.2

Esta é uma versão de patch para adicionar a autenticação AWS Signature versão 4 para solicitações ao Amazon S3. Todos os aplicativos e componentes são iguais aos da EMR versão anterior da Amazon.

Important

Nesta versão de lançamento, a Amazon EMR usa o AWS Signature versão 4 exclusivamente para autenticar solicitações para o Amazon S3. Para obter mais informações, consulte [Novidades](#).

Versões de componente 5.14.2

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR diferem das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.0.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.5.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.4.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.10.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
<code>emrfs</code>	2.23.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
flink-client	1.4.2	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.3-amzn-1	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.3-amzn-1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.8.3-amzn-1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.8.3-amzn-1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httpfs-server	2.8.3-amzn-1	HTTPendpoint para HDFS operações.

Componente	Version (Versão)	Descrição
hadoop-kms-server	2.8.3-amzn-1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.8.3-amzn-1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.8.3-amzn-1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.8.3-amzn-1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.3-amzn-1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.2	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.2	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.4.2	HBasecliente de linha de comando.
hbase-rest-server	1.4.2	Serviço que fornece um RESTful HTTP endpoint paraHBase.

Componente	Version (Versão)	Descrição
hbase-thrift-server	1.4.2	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	2.3.2-amzn-2	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.2-amzn-2	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.2-amzn-2	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	2.3.2-amzn-2	O cliente da linha de comando do Hive.
hive-hbase	2.3.2-amzn-2	Cliente do Hive-hbase.
hive-metastore-server	2.3.2-amzn-2	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.2-amzn-2	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.1.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop

Componente	Version (Versão)	Descrição
jupyterhub	0.8.1	Servidor de vários usuários para blocos de anotações Jupyter
livy-server	0.4.0-incubating	RESTinterface para interagir com o Apache Spark
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.1.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.1.85	Drivers NVIDIA e toolkit CUDA
oozie-client	4.3.0	O cliente da linha de comando do Oozie.
oozie-server	4.3.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	3.4.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.13.0- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.13.0- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API

Componente	Version (Versão)	Descrição
presto-coordinator	0.194	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.194	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.1	O projeto R para computação estatística
spark-client	2.3.0	Os clientes da linha de comando do Spark.
spark-history-server	2.3.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.3.0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.3.0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.

Componente	Version (Versão)	Descrição
zeppelin-server	0.7.3	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.14.2

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.14.2

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .

Classificações	Descrição
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.

Classificações	Descrição
hdfs-encryption-zones	Configure as zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.

Classificações	Descrição
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente do HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.

Classificações	Descrição
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.

Classificações	Descrição
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.

Classificações	Descrição
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.

Classificações	Descrição
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.14.1

Versões 5.14.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.14.1	emr-5.14.0	emr-5.13.1	emr-5.13.0
AWS SDK para 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

Notas da versão 5.14.1

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.14.1 da Amazon. As alterações são referentes à versão 5.14.0.

Data da versão inicial: 17 de outubro de 2018

Atualizou o padrão da Amazon AMI EMR para resolver possíveis vulnerabilidades de segurança.

Versões de componente 5.14.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou

aws. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR diferem das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado myapp-component com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como 2.2-amzn-2.

Componente	Version (Versão)	Descrição
aws-sagemaker-spark-sdk	1.0.1	Amazon SageMaker Spark SDK
emr-ddb	4.5.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.4.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.10.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.23.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.4.2	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do

Componente	Version (Versão)	Descrição
		ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.3-amzn-1	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.3-amzn-1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.8.3-amzn-1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.8.3-amzn-1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httpfs-server	2.8.3-amzn-1	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.8.3-amzn-1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API

Componente	Version (Versão)	Descrição
hadoop-mapred	2.8.3-amzn-1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.8.3-amzn-1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.8.3-amzn-1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.3-amzn-1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.2	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.2	Serviço para atender uma ou mais HBase regiões.
hbase-client	1.4.2	HBasecliente de linha de comando.
hbase-rest-server	1.4.2	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	1.4.2	Serviço que fornece um endpoint Thrift para. HBase

Componente	Version (Versão)	Descrição
hcatalog-client	2.3.2-amzn-2	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.2-amzn-2	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.2-amzn-2	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	2.3.2-amzn-2	O cliente da linha de comando do Hive.
hive-hbase	2.3.2-amzn-2	Cliente do Hive-hbase.
hive-metastore-server	2.3.2-amzn-2	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.2-amzn-2	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.1.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	0.8.1	Servidor de vários usuários para blocos de anotações Jupyter

Componente	Version (Versão)	Descrição
livy-server	0.4.0-incubating	REST interface para interagir com o Apache Spark
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.1.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.1.85	Drivers NVIDIA e toolkit CUDA
oozie-client	4.3.0	O cliente da linha de comando do Oozie.
oozie-server	4.3.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	3.4.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.13.0- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.13.0- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.194	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.

Componente	Version (Versão)	Descrição
presto-worker	0.194	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.1	O projeto R para computação estatística
spark-client	2.3.0	Os clientes da linha de comando do Spark.
spark-history-server	2.3.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.3.0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.3.0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.
zeppelin-server	0.7.3	O notebook baseado na web que permite a análise de dados interativa.

Componente	Version (Versão)	Descrição
zookeeper-server	3.4.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.14.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.14.1

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.

Classificações	Descrição
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure as zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.

Classificações	Descrição
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.

Classificações	Descrição
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.

Classificações	Descrição
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.

Classificações	Descrição
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.

Classificações	Descrição
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oracoop-site	Altere os valores no arquivo oracoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.

Classificações	Descrição
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.14.0

Versões 5.14.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [JupyterHub](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.14.0	emr-5.13.1	emr-5.13.0	emr-5.12.3
AWS SDK para 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

	emr-5.14.0	emr-5.13.1	emr-5.13.0	emr-5.12.3
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
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

	emr-5.14.0	emr-5.13.1	emr-5.13.0	emr-5.12.3
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

Notas da versão 5.14.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.14.0 da Amazon. As alterações são referentes à versão 5.13.0.

Data da versão inicial: 4 de junho de 2018

Atualizações

- Atualizado Apache Flink para 1.4.2
- Apache MXnet atualizado para 1.1.0
- Atualizado Apache Sqoop para 1.4.7

Novos atributos

- JupyterHub Suporte adicionado. Para ter mais informações, consulte [JupyterHub](#).

Alterações, melhorias e problemas resolvidos

- EMRFS
 - A userAgent sequência de caracteres nas solicitações para o Amazon S3 foi atualizada para conter as informações de usuário e grupo do principal invocador. Isso pode ser usado com AWS CloudTrail registros para um rastreamento de solicitações mais abrangente.
- HBase
 - Inclui [HBASE-20447](#), que soluciona um problema que pode causar problemas de cache, especialmente com regiões divididas.
- MXnet
 - Adicionadas bibliotecas OpenCV.
- Spark
 - Quando o Spark grava arquivos Parquet em um local do Amazon S3 EMRFS usando, FileOutputCommitter o algoritmo foi atualizado para usar a versão 2 em vez da versão 1. Isso reduz o número de renomeações, o que melhora a performance da aplicação. Essa alteração não afeta:
 - Aplicações diferentes do Spark.
 - Aplicativos que gravam em outros sistemas de arquivos, como HDFS (que ainda usam a versão 1 do FileOutputCommitter).
 - Aplicativos que usam outros formatos de saída, como texto ou csv, que já usam gravação EMRFS direta.

Problemas conhecidos

- JupyterHub
 - O uso de classificações de configuração para configurar JupyterHub notebooks Jupyter individuais ao criar um cluster não é suportado. Edite manualmente o arquivo jupyterhub_config.py e os arquivos jupyter_notebook_config.py para cada usuário. Para ter mais informações, consulte [Configurando JupyterHub](#).
 - JupyterHub falha ao iniciar em clusters dentro de uma sub-rede privada, falhando com a mensagem. `Error: ENOENT: no such file or directory, open '/etc/jupyter/conf/server.crt'` Isso é causado por um erro no script que gera certificados autoassinados. Use a solução alternativa a seguir para gerar certificados autoassinados. Todos os comandos são executados enquanto estão conectados ao nó primário.

1. Copie o script de geração de certificados do contêiner para o nó primário:

```
sudo docker cp jupyterhub:/tmp/gen_self_signed_cert.sh ./
```

2. Use um editor de texto para alterar a linha 23 e mudar o nome de host público para o nome de host local, conforme mostrado abaixo:

```
local hostname=$(curl -s $EC2_METADATA_SERVICE_URI/local-hostname)
```

3. Execute o script para gerar certificados autoassinados:

```
sudo bash ./gen_self_signed_cert.sh
```

4. Mova os arquivos de certificado que o script gera para o diretório `/etc/jupyter/conf/`:

```
sudo mv /tmp/server.crt /tmp/server.key /etc/jupyter/conf/
```

Você pode acessar `tail` o `jupyter.log` arquivo para verificar se ele JupyterHub foi reiniciado e está retornando um código de resposta 200. Por exemplo: .

```
tail -f /var/log/jupyter/jupyter.log
```

Essa ação deve retornar uma resposta semelhante à seguinte:

```
# [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
```

- Depois que o nó primário for reinicializado ou o controlador de instância for reiniciado, as CloudWatch métricas não serão coletadas e o recurso de escalabilidade automática não estará disponível nas EMR versões 5.14.0, 5.15.0 ou 5.16.0 da Amazon. Esse problema foi corrigido no Amazon EMR 5.17.0.

Versões de componente 5.14.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são

a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR diferem das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado myapp-component com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como 2.2-amzn-2.

Componente	Version (Versão)	Descrição
aws-sagemaker-spark-sdk	1.0.1	Amazon SageMaker Spark SDK
emr-ddb	4.5.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.4.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.10.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.23.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.4.2	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop,

Componente	Version (Versão)	Descrição
		juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.3-amzn-1	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.3-amzn-1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.8.3-amzn-1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.8.3-amzn-1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httpfs-server	2.8.3-amzn-1	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.8.3-amzn-1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API

Componente	Version (Versão)	Descrição
hadoop-mapred	2.8.3-amzn-1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.8.3-amzn-1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.8.3-amzn-1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.3-amzn-1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.2	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.2	Serviço para atender uma ou mais HBase regiões.
hbase-client	1.4.2	HBasecliente de linha de comando.
hbase-rest-server	1.4.2	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	1.4.2	Serviço que fornece um endpoint Thrift para. HBase

Componente	Version (Versão)	Descrição
hcatalog-client	2.3.2-amzn-2	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.2-amzn-2	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.2-amzn-2	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	2.3.2-amzn-2	O cliente da linha de comando do Hive.
hive-hbase	2.3.2-amzn-2	Cliente do Hive-hbase.
hive-metastore-server	2.3.2-amzn-2	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.2-amzn-2	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.1.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
jupyterhub	0.8.1	Servidor de vários usuários para blocos de anotações Jupyter

Componente	Version (Versão)	Descrição
livy-server	0.4.0-incubating	RESTinterface para interagir com o Apache Spark
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.1.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.1.85	Drivers NVIDIA e toolkit CUDA
oozie-client	4.3.0	O cliente da linha de comando do Oozie.
oozie-server	4.3.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
opencv	3.4.0	Biblioteca de Visão Computacional de Código Aberto.
phoenix-library	4.13.0- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.13.0- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.194	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.

Componente	Version (Versão)	Descrição
presto-worker	0.194	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.1	O projeto R para computação estatística
spark-client	2.3.0	Os clientes da linha de comando do Spark.
spark-history-server	2.3.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.3.0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.3.0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.7	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.
zeppelin-server	0.7.3	O notebook baseado na web que permite a análise de dados interativa.

Componente	Version (Versão)	Descrição
zookeeper-server	3.4.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.14.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.14.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
container-log4j	Altere os valores no arquivo <code>container-log4j.properties</code> YARN do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.

Classificações	Descrição
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure as zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.

Classificações	Descrição
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.

Classificações	Descrição
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
jupyter-notebook-conf	Alterar os valores no arquivo jupyter_notebook_config.py do Notebook Jupyter.
jupyter-hub-conf	Altere os valores no JupyterHubs arquivo jupyterhub_config.py do.
jupyter-sparkmagic-conf	Altere os valores no arquivo config.json do Sparkmagic.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.

Classificações	Descrição
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.

Classificações	Descrição
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.

Classificações	Descrição
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oracoop-site	Altere os valores no arquivo oracoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.

Classificações	Descrição
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.13.1

Versões 5.13.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.13.1	emr-5.13.0	emr-5.12.3	emr-5.12.2
AWS SDK para 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

	emr-5.13.1	emr-5.13.0	emr-5.12.3	emr-5.12.2
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	-	-	-	-
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

	emr-5.13.1	emr-5.13.0	emr-5.12.3	emr-5.12.2
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

Notas da versão 5.13.1

Esta é uma versão de patch para adicionar a autenticação AWS Signature versão 4 para solicitações ao Amazon S3. Todos os aplicativos e componentes são iguais aos da EMR versão anterior da Amazon.

Important

Nesta versão de lançamento, a Amazon EMR usa o AWS Signature versão 4 exclusivamente para autenticar solicitações para o Amazon S3. Para obter mais informações, consulte [Novidades](#).

Versões de componente 5.13.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou

aws. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR diferem das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado myapp-component com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como 2.2-amzn-2.

Componente	Version (Versão)	Descrição
aws-sagemaker-spark-sdk	1.0.1	Amazon SageMaker Spark SDK
emr-ddb	4.5.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.4.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.10.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.22.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.4.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do

Componente	Version (Versão)	Descrição
		ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.3-amzn-0	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.3-amzn-0	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.8.3-amzn-0	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.8.3-amzn-0	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-https-server	2.8.3-amzn-0	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.8.3-amzn-0	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API

Componente	Version (Versão)	Descrição
hadoop-mapred	2.8.3-amzn-0	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.8.3-amzn-0	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.8.3-amzn-0	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.3-amzn-0	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.2	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.2	Serviço para atender uma ou mais HBase regiões.
hbase-client	1.4.2	HBasecliente de linha de comando.
hbase-rest-server	1.4.2	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	1.4.2	Serviço que fornece um endpoint Thrift para. HBase

Componente	Version (Versão)	Descrição
hcatalog-client	2.3.2-amzn-2	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.2-amzn-2	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.2-amzn-2	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	2.3.2-amzn-2	O cliente da linha de comando do Hive.
hive-hbase	2.3.2-amzn-2	Cliente do Hive-hbase.
hive-metastore-server	2.3.2-amzn-2	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.2-amzn-2	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.1.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
livy-server	0.4.0-incubating	RESTinterface para interagir com o Apache Spark

Componente	Version (Versão)	Descrição
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.0.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.1.85	Drivers NVIDIA e toolkit CUDA
oozie-client	4.3.0	O cliente da linha de comando do Oozie.
oozie-server	4.3.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.13.0- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.13.0- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.194	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.194	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.

Componente	Version (Versão)	Descrição
r	3.4.1	O projeto R para computação estatística
spark-client	2.3.0	Os clientes da linha de comando do Spark.
spark-history-server	2.3.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.3.0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.3.0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.
zeppelin-server	0.7.3	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.

Componente	Version (Versão)	Descrição
zookeeper-client	3.4.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.13.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.13.1

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Alterar as configurações de <code>log4j-cli.properties</code> no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Hadoop.

Classificações	Descrição
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure as zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.

Classificações	Descrição
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.

Classificações	Descrição
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.

Classificações	Descrição
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.

Classificações	Descrição
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.

Classificações	Descrição
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.13.0

Versões 5.13.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.13.0	emr-5.12.3	emr-5.12.2	emr-5.12.1
AWS SDK para 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

	emr-5.13.0	emr-5.12.3	emr-5.12.2	emr-5.12.1
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	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	0.188	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

Notas da versão 5.13.0

As notas de lançamento a seguir incluem informações para a EMR versão 5.13.0 da Amazon. As alterações são referentes à versão 5.12.0.

Atualizações

- Atualizado Spark para 2.3.0
- Atualizado HBase para 1.4.2
- Atualizado Presto para 0.194
- Atualizado para AWS SDK for Java 1.11.297

Alterações, melhorias e problemas resolvidos

- Hive
 - Portado para trás [HIVE-15436](#). Hive aprimorado APIs para retornar somente visualizações.

Problemas conhecidos

- MXNet atualmente não tem bibliotecas OpenCV.

Versões de componente 5.13.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR diferem das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
aws-sagemaker-spark-sdk	1.0.1	Amazon SageMaker Spark SDK
emr-ddb	4.5.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.4.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.10.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.22.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.4.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.

Componente	Version (Versão)	Descrição
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.3-amzn-0	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.3-amzn-0	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.8.3-amzn-0	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.8.3-amzn-0	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-https-server	2.8.3-amzn-0	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.8.3-amzn-0	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.8.3-amzn-0	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.8.3-amzn-0	YARNserviço para gerenciar contêineres em um nó individual.

Componente	Version (Versão)	Descrição
hadoop-yarn-resourcemanager	2.8.3-amzn-0	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.3-amzn-0	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.2	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.2	Serviço para atender uma ou mais HBase regiões.
hbase-client	1.4.2	HBasecliente de linha de comando.
hbase-rest-server	1.4.2	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	1.4.2	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	2.3.2-amzn-2	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.2-amzn-2	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.

Componente	Version (Versão)	Descrição
hcatalog-webhcat-server	2.3.2-amzn-2	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	2.3.2-amzn-2	O cliente da linha de comando do Hive.
hive-hbase	2.3.2-amzn-2	Cliente do Hive-hbase.
hive-metastore-server	2.3.2-amzn-2	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.2-amzn-2	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.1.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
livy-server	0.4.0-incubating	RESTinterface para interagir com o Apache Spark
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.0.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.1.85	Drivers NVIDIA e toolkit CUDA

Componente	Version (Versão)	Descrição
oozie-client	4.3.0	O cliente da linha de comando do Oozie.
oozie-server	4.3.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.13.0- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.13.0- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.194	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.194	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
r	3.4.1	O projeto R para computação estatística
spark-client	2.3.0	Os clientes da linha de comando do Spark.
spark-history-server	2.3.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.

Componente	Version (Versão)	Descrição
spark-on-yarn	2.3.0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.3.0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.
zeppelin-server	0.7.3	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.13.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.13.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.

Classificações	Descrição
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure as zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.

Classificações	Descrição
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.

Classificações	Descrição
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.

Classificações	Descrição
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.

Classificações	Descrição
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.

Classificações	Descrição
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.12.3

Versões 5.12.3 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.12.3	emr-5.12.2	emr-5.12.1	emr-5.12.0
AWS SDK para 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

Notas da versão 5.12.3

Esta é uma versão de patch para adicionar a autenticação AWS Signature versão 4 para solicitações ao Amazon S3. Todos os aplicativos e componentes são iguais aos da EMR versão anterior da Amazon.

Important

Nesta versão de lançamento, a Amazon EMR usa o AWS Signature versão 4 exclusivamente para autenticar solicitações para o Amazon S3. Para obter mais informações, consulte [Novidades](#).

Versões de componente 5.12.3

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR diferem das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.0.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.5.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.4.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.9.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
<code>emrfs</code>	2.21.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
flink-client	1.4.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.3-amzn-0	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.3-amzn-0	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.8.3-amzn-0	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.8.3-amzn-0	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httpfs-server	2.8.3-amzn-0	HTTPendpoint para HDFS operações.

Componente	Version (Versão)	Descrição
hadoop-kms-server	2.8.3-amzn-0	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.8.3-amzn-0	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.8.3-amzn-0	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.8.3-amzn-0	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.3-amzn-0	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.0	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.0	Serviço para atender uma ou mais HBase regiões.
hbase-client	1.4.0	HBasecliente de linha de comando.
hbase-rest-server	1.4.0	Serviço que fornece um RESTful HTTP endpoint paraHBase.

Componente	Version (Versão)	Descrição
hbase-thrift-server	1.4.0	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	2.3.2-amzn-1	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.2-amzn-1	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.2-amzn-1	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	2.3.2-amzn-1	O cliente da linha de comando do Hive.
hive-hbase	2.3.2-amzn-1	Cliente do Hive-hbase.
hive-metastore-server	2.3.2-amzn-1	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.2-amzn-1	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.1.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop

Componente	Version (Versão)	Descrição
livy-server	0.4.0-incubating	REST interface para interagir com o Apache Spark
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.0.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.1.85	Drivers NVIDIA e toolkit CUDA
oozie-client	4.3.0	O cliente da linha de comando do Oozie.
oozie-server	4.3.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.13.0- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.13.0- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.188	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.188	O serviço que executa partes de uma consulta.

Componente	Version (Versão)	Descrição
pig-client	0.17.0	O cliente da linha de comando do Pig.
spark-client	2.2.1	Os clientes da linha de comando do Spark.
spark-history-server	2.2.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.2.1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.2.1	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.
zeppelin-server	0.7.3	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.

Componente	Version (Versão)	Descrição
zookeeper-client	3.4.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.12.3

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.12.3

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Alterar as configurações de <code>log4j-cli.properties</code> no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Hadoop.

Classificações	Descrição
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure as zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.

Classificações	Descrição
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.

Classificações	Descrição
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.

Classificações	Descrição
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.

Classificações	Descrição
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.

Classificações	Descrição
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.12.2

Versões 5.12.2 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.12.2	emr-5.12.1	emr-5.12.0	emr-5.11.4
AWS SDK para 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	-	-	-	-
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

	emr-5.12.2	emr-5.12.1	emr-5.12.0	emr-5.11.4
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
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

Notas da versão 5.12.2

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.12.2 da Amazon. As alterações são referentes à versão 5.12.1.

Data da versão inicial: 29 de agosto de 2018

Alterações, melhorias e problemas resolvidos

- Esta versão aborda uma possível vulnerabilidade de segurança.

Versões de componente 5.12.2

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.0.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.5.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.4.0	Bibliotecas convenientes para o ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
emr-kinesis	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.9.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.21.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.4.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.3-amzn-0	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.3-amzn-0	HDFSserviço em nível de nó para armazenar blocos.

Componente	Version (Versão)	Descrição
hadoop-hdfs-library	2.8.3-amzn-0	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.8.3-amzn-0	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httfs-server	2.8.3-amzn-0	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.8.3-amzn-0	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.8.3-amzn-0	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.8.3-amzn-0	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.8.3-amzn-0	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.3-amzn-0	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.0	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.

Componente	Version (Versão)	Descrição
hbase-region-server	1.4.0	Serviço para atender uma ou mais HBase regiões.
hbase-client	1.4.0	HBase cliente de linha de comando.
hbase-rest-server	1.4.0	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.4.0	Serviço que fornece um endpoint Thrift para HBase
hcatalog-client	2.3.2-amzn-1	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.2-amzn-1	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.2-amzn-1	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.3.2-amzn-1	O cliente da linha de comando do Hive.
hive-hbase	2.3.2-amzn-1	Cliente do Hive-hbase.
hive-metastore-server	2.3.2-amzn-1	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL

Componente	Version (Versão)	Descrição
hive-server2	2.3.2-amzn-1	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.1.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
livy-server	0.4.0-incubating	RESTinterface para interagir com o Apache Spark
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.0.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.1.85	Drivers NVIDIA e toolkit CUDA
oozie-client	4.3.0	O cliente da linha de comando do Oozie.
oozie-server	4.3.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.13.0- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.13.0- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API

Componente	Version (Versão)	Descrição
presto-coordinator	0.188	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.188	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
spark-client	2.2.1	Os clientes da linha de comando do Spark.
spark-history-server	2.2.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.2.1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.2.1	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.

Componente	Version (Versão)	Descrição
zeppelin-server	0.7.3	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.12.2

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.12.2

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.

Classificações	Descrição
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure as zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.

Classificações	Descrição
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.

Classificações	Descrição
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.

Classificações	Descrição
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.

Classificações	Descrição
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.

Classificações	Descrição
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.12.1

Versões 5.12.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.12.1	emr-5.12.0	emr-5.11.4	emr-5.11.3
AWS SDK para 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

	emr-5.12.1	emr-5.12.0	emr-5.11.4	emr-5.11.3
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	-	-	-	-
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

	emr-5.12.1	emr-5.12.0	emr-5.11.4	emr-5.11.3
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

Notas da versão 5.12.1

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.12.1 da Amazon. As alterações são referentes à versão 5.12.0.

Data da versão inicial: 29 de março de 2018

Alterações, melhorias e problemas resolvidos

- Atualizou o kernel Amazon Linux do default Amazon Linux AMI para Amazon EMR para resolver possíveis vulnerabilidades.

Versões de componente 5.12.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
aws-sagemaker-spark-sdk	1.0.1	Amazon SageMaker Spark SDK
emr-ddb	4.5.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.4.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.9.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.21.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.4.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.

Componente	Version (Versão)	Descrição
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.3-amzn-0	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.8.3-amzn-0	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.8.3-amzn-0	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.8.3-amzn-0	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-https-server	2.8.3-amzn-0	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.8.3-amzn-0	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.8.3-amzn-0	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.8.3-amzn-0	YARNserviço para gerenciar contêineres em um nó individual.

Componente	Version (Versão)	Descrição
hadoop-yarn-resourcemanager	2.8.3-amzn-0	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.3-amzn-0	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.4.0	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.0	Serviço para atender uma ou mais HBase regiões.
hbase-client	1.4.0	HBasecliente de linha de comando.
hbase-rest-server	1.4.0	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	1.4.0	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	2.3.2-amzn-1	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.2-amzn-1	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.

Componente	Version (Versão)	Descrição
hcatalog-webhcat-server	2.3.2-amzn-1	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	2.3.2-amzn-1	O cliente da linha de comando do Hive.
hive-hbase	2.3.2-amzn-1	Cliente do Hive-hbase.
hive-metastore-server	2.3.2-amzn-1	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.2-amzn-1	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.1.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
livy-server	0.4.0-incubating	RESTinterface para interagir com o Apache Spark
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.0.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.1.85	Drivers NVIDIA e toolkit CUDA

Componente	Version (Versão)	Descrição
oozie-client	4.3.0	O cliente da linha de comando do Oozie.
oozie-server	4.3.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.13.0- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.13.0- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.188	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.188	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
spark-client	2.2.1	Os clientes da linha de comando do Spark.
spark-history-server	2.2.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.2.1	Mecanismo de execução na memória para YARN.

Componente	Version (Versão)	Descrição
spark-yarn-slave	2.2.1	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.7.3	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.12.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.12.1

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.

Classificações	Descrição
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure as zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.

Classificações	Descrição
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.

Classificações	Descrição
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.

Classificações	Descrição
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.

Classificações	Descrição
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.

Classificações	Descrição
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.12.0

Versões da aplicação 5.12.0

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.12.0	emr-5.11.4	emr-5.11.3	emr-5.11.2
AWS SDK para 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

Notas da versão 5.12.0

As notas de lançamento a seguir incluem informações para a EMR versão 5.12.0 da Amazon. As alterações são referentes à versão 5.11.1.

Atualizações

- AWS SDK para Java 1.11.238 ⇒ 1.11.267. Para obter mais informações, consulte [AWS SDKo log de alterações do Java](#) GitHub.
- Hadoop 2.7.3 ⇒ 2.8.3. Para obter mais informações, consulte [Versões do Apache Hadoop](#).
- Flink 1.3.2 ⇒ 1.4.0. Para obter mais informações, consulte o [Anúncio de versão do Apache Flink 1.4.0](#).
- HBase 1.3.1 ⇒ 1.4.0. Para obter mais informações, consulte o [anúncio HBase de lançamento](#).

- Hue 4.0.1 ⇒ 4.1.0. Para obter mais informações, consulte as [Notas de versão](#).
- MxNet 0,12,0 ⇒ 1,0,0. Para obter mais informações, consulte o [MXNetChange Log](#) on GitHub.
- Presto 0.187 ⇒ 0.188. Para obter mais informações, consulte as [Notas de versão](#).

Alterações, melhorias e problemas resolvidos

- Hadoop
 - A propriedade `yarn.resourcemanager.decommissioning.timeout` foi alterada para `yarn.resourcemanager.nodemanager-graceful-decommission-timeout-secs`. Você pode usar essa propriedade para personalizar a redução da escala do cluster verticalmente. Para obter mais informações, consulte [Cluster Scale-Down](#) no Amazon EMR Management Guide.
 - O Hadoop CLI adicionou a `-d` opção ao comando `cp` (copiar), que especifica a cópia direta. Você pode usar isso para evitar a criação de um arquivo `.COPYING` intermediário, o que torna mais rápida a cópia de dados entre o Amazon S3. Para obter mais informações, consulte [HADOOP-12384](#).
- Pig
 - Adicionada a classificação de configuração `pig-env`, que simplifica a configuração das propriedades do ambiente do Pig. Para ter mais informações, consulte [Configurar aplicações](#).
- Presto
 - Adicionada a classificação de configuração `presto-connector-redshift`, que pode ser usada para configurar os valores no arquivo de configuração `redshift.properties` do Presto. Para obter mais informações, consulte [Conector do Redshift](#) na documentação do Presto e [Configurar aplicações](#).
 - O suporte do Presto para EMRFS foi adicionado e é a configuração padrão. As EMR versões anteriores da Amazon usavam o `PrestOS3FileSystem`, que era a única opção. Para ter mais informações, consulte [EMRFS e configuração do PrestOS3 FileSystem](#).

Note

Se você consultar dados subjacentes no Amazon S3 com a EMR versão 5.12.0 da Amazon, podem ocorrer erros do Presto. Isso acontece porque o Presto não consegue obter valores de classificação de configuração em `emrfs-site.xml`. Como solução alternativa, crie um subdiretório `emrfs` em `usr/lib/presto/plugin/hive-hadoop2/` e crie um link simbólico em `usr/lib/presto/plugin/hive-hadoop2/`

`emrfs` para o arquivo `/usr/share/aws/emr/emrfs/conf/emrfs-site.xml` existente. Em seguida, reinicie o processo `presto-server` (`sudo presto-server stop` seguido por `sudo presto-server start`).

- Spark
 - Backported [SPARK-22036: a BigDecimal multiplicação](#) às vezes retorna nula.

Problemas conhecidos

- MXNet não inclui bibliotecas OpenCV.
- O SparkR não está disponível para clusters criados usando um customizado porque o AMI R não é instalado por padrão nos nós do cluster.

Versões de componente 5.12.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1.0.1	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.5.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
emr-goodies	2.4.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.9.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.21.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.4.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.8.3-amzn-0	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.

Componente	Version (Versão)	Descrição
hadoop-hdfs-datanode	2.8.3-amzn-0	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.8.3-amzn-0	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.8.3-amzn-0	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httfs-server	2.8.3-amzn-0	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.8.3-amzn-0	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.8.3-amzn-0	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.8.3-amzn-0	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.8.3-amzn-0	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.8.3-amzn-0	Serviço para recuperar informações atuais e históricas para YARN aplicativos.

Componente	Version (Versão)	Descrição
hbase-hmaster	1.4.0	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.4.0	Serviço para atender uma ou mais HBase regiões.
hbase-client	1.4.0	HBase cliente de linha de comando.
hbase-rest-server	1.4.0	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.4.0	Serviço que fornece um endpoint Thrift para HBase.
hcatalog-client	2.3.2-amzn-1	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.2-amzn-1	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.2-amzn-1	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.3.2-amzn-1	O cliente da linha de comando do Hive.
hive-hbase	2.3.2-amzn-1	Cliente do Hive-hbase.

Componente	Version (Versão)	Descrição
hive-metastore-server	2.3.2-amzn-1	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.2-amzn-1	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.1.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
livy-server	0.4.0-incubating	RESTinterface para interagir com o Apache Spark
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	1.0.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.1.85	Drivers NVIDIA e toolkit CUDA
oozie-client	4.3.0	O cliente da linha de comando do Oozie.
oozie-server	4.3.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.

Componente	Version (Versão)	Descrição
phoenix-library	4.13.0- -1,4 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.13.0- -1,4 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.188	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.188	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
spark-client	2.2.1	Os clientes da linha de comando do Spark.
spark-history-server	2.2.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.2.1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.2.1	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.

Componente	Version (Versão)	Descrição
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.7.3	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.12.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.12.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.

Classificações	Descrição
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.

Classificações	Descrição
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure as zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.

Classificações	Descrição
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.

Classificações	Descrição
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-env	Alterar os valores no ambiente do Pig.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.

Classificações	Descrição
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.

Classificações	Descrição
presto-connector-redshift	Alterar os valores no arquivo de propriedades redshift do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.

Classificações	Descrição
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.11.4

Versões 5.11.4 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.11.4	emr-5.11.3	emr-5.11.2	emr-5.11.1
AWS SDK para 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

	emr-5.11.4	emr-5.11.3	emr-5.11.2	emr-5.11.1
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
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

	emr-5.11.4	emr-5.11.3	emr-5.11.2	emr-5.11.1
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

Notas da versão 5.11.4

Esta é uma versão de patch para adicionar a autenticação AWS Signature versão 4 para solicitações ao Amazon S3. Todos os aplicativos e componentes são iguais aos da EMR versão anterior da Amazon.

Important

Nesta versão de lançamento, a Amazon EMR usa o AWS Signature versão 4 exclusivamente para autenticar solicitações para o Amazon S3. Para obter mais informações, consulte [Novidades](#).

Versões de componente 5.11.4

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com emr ou

aws. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado myapp-component com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como 2.2-amzn-2.

Componente	Version (Versão)	Descrição
aws-sagemaker-spark-sdk	1,0	Amazon SageMaker Spark SDK
emr-ddb	4.5.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.4.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.8.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.20.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.3.2	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do

Componente	Version (Versão)	Descrição
		ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-6	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-6	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-6	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-6	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-https-server	2.7.3-amzn-6	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-6	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API

Componente	Version (Versão)	Descrição
hadoop-mapred	2.7.3-amzn-6	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-6	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-6	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-6	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.3.1	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.3.1	Serviço para atender uma ou mais HBase regiões.
hbase-client	1.3.1	HBasecliente de linha de comando.
hbase-rest-server	1.3.1	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	1.3.1	Serviço que fornece um endpoint Thrift para. HBase

Componente	Version (Versão)	Descrição
hcatalog-client	2.3.2-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.2-amzn-0	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.2-amzn-0	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	2.3.2-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.3.2-amzn-0	Cliente do Hive-hbase.
hive-metastore-server	2.3.2-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.2-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.0.1	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
livy-server	0.4.0-incubating	RESTinterface para interagir com o Apache Spark

Componente	Version (Versão)	Descrição
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	0.12.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.0.176	Drivers NVIDIA e toolkit CUDA
oozie-client	4.3.0	O cliente da linha de comando do Oozie.
oozie-server	4.3.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.11.0- -1.3 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.11.0- -1.3 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.187	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.187	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.

Componente	Version (Versão)	Descrição
spark-client	2.2.1	Os clientes da linha de comando do Spark.
spark-history-server	2.2.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.2.1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.2.1	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.
zeppelin-server	0.7.3	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.11.4

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.11.4

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Alterar as configurações de <code>log4j-cli.properties</code> no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop

Classificações	Descrição
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure as zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat

Classificações	Descrição
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.

Classificações	Descrição
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.

Classificações	Descrição
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.

Classificações	Descrição
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.

Classificações	Descrição
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.11.3

Versões 5.11.3 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.11.3	emr-5.11.2	emr-5.11.1	emr-5.11.0
AWS SDK para 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-5.11.3	emr-5.11.2	emr-5.11.1	emr-5.11.0
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
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

Notas da versão 5.11.3

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.11.3 da Amazon. As alterações são referentes à versão 5.11.2.

Data da versão inicial: 18 de julho de 2019

Alterações, melhorias e problemas resolvidos

- Atualizou o Amazon Linux padrão AMI para Amazon EMR para incluir atualizações importantes de segurança do kernel Linux, incluindo o problema de TCP SACK negação de serviço ([AWS-2019-005](#)).

Versões de componente 5.11.3

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1,0	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.5.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.4.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.8.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
<code>emrfs</code>	2.20.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
flink-client	1.3.2	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-6	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-6	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-6	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-6	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httpfs-server	2.7.3-amzn-6	HTTPendpoint para HDFS operações.

Componente	Version (Versão)	Descrição
hadoop-kms-server	2.7.3-amzn-6	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-6	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-6	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-6	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-6	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.3.1	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.3.1	Serviço para atender uma ou mais HBase regiões.
hbase-client	1.3.1	HBasecliente de linha de comando.
hbase-rest-server	1.3.1	Serviço que fornece um RESTful HTTP endpoint paraHBase.

Componente	Version (Versão)	Descrição
hbase-thrift-server	1.3.1	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	2.3.2-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.2-amzn-0	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.2-amzn-0	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	2.3.2-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.3.2-amzn-0	Cliente do Hive-hbase.
hive-metastore-server	2.3.2-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.2-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.0.1	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop

Componente	Version (Versão)	Descrição
livy-server	0.4.0-incubating	REST interface para interagir com o Apache Spark
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	0.12.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.0.176	Drivers NVIDIA e toolkit CUDA
oozie-client	4.3.0	O cliente da linha de comando do Oozie.
oozie-server	4.3.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.11.0- -1.3 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.11.0- -1.3 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.187	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.187	O serviço que executa partes de uma consulta.

Componente	Version (Versão)	Descrição
pig-client	0.17.0	O cliente da linha de comando do Pig.
spark-client	2.2.1	Os clientes da linha de comando do Spark.
spark-history-server	2.2.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.2.1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.2.1	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.
zeppelin-server	0.7.3	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.

Componente	Version (Versão)	Descrição
zookeeper-client	3.4.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.11.3

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.11.3

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Alterar as configurações de <code>log4j-cli.properties</code> no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Hadoop.

Classificações	Descrição
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure as zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.

Classificações	Descrição
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.

Classificações	Descrição
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.

Classificações	Descrição
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.

Classificações	Descrição
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.

Classificações	Descrição
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.11.2

Versões 5.11.2 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)

- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.11.2	emr-5.11.1	emr-5.11.0	emr-5.10.1
AWS SDK para 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	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-

	emr-5.11.2	emr-5.11.1	emr-5.11.0	emr-5.10.1
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

Notas da versão 5.11.2

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.11.2 da Amazon. As alterações são referentes à versão 5.11.1.

Data da versão inicial: 29 de agosto de 2018

Alterações, melhorias e problemas resolvidos

- Esta versão aborda uma possível vulnerabilidade de segurança.

Versões de componente 5.11.2

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1,0	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.5.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.4.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.8.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.

Componente	Version (Versão)	Descrição
emrfs	2.20.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.3.2	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-6	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-6	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-6	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-6	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.

Componente	Version (Versão)	Descrição
hadoop-https-server	2.7.3-amzn-6	HTTP endpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-6	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-6	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-6	YARN serviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-6	YARN serviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-6	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.3.1	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.3.1	Serviço para atender uma ou mais HBase regiões.
hbase-client	1.3.1	HBase cliente de linha de comando.

Componente	Version (Versão)	Descrição
hbase-rest-server	1.3.1	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.3.1	Serviço que fornece um endpoint Thrift para HBase
hcatalog-client	2.3.2-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.2-amzn-0	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.2-amzn-0	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.3.2-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.3.2-amzn-0	Cliente do Hive-hbase.
hive-metastore-server	2.3.2-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.2-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.

Componente	Version (Versão)	Descrição
hue-server	4.0.1	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
livy-server	0.4.0-incubating	RESTinterface para interagir com o Apache Spark
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	0.12.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.0.176	Drivers NVIDIA e toolkit CUDA
oozie-client	4.3.0	O cliente da linha de comando do Oozie.
oozie-server	4.3.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.11.0- -1.3 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.11.0- -1.3 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API

Componente	Version (Versão)	Descrição
presto-coordinator	0.187	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.187	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
spark-client	2.2.1	Os clientes da linha de comando do Spark.
spark-history-server	2.2.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.2.1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.2.1	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.

Componente	Version (Versão)	Descrição
zeppelin-server	0.7.3	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.11.2

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.11.2

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.

Classificações	Descrição
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure as zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.

Classificações	Descrição
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.

Classificações	Descrição
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.

Classificações	Descrição
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.

Classificações	Descrição
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.

Classificações	Descrição
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.11.1

Versões 5.11.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.11.1	emr-5.11.0	emr-5.10.1	emr-5.10.0
AWS SDK para 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
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	-	-	-	-

	emr-5.11.1	emr-5.11.0	emr-5.10.1	emr-5.10.0
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	2.2.0	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

Notas da versão 5.11.1

As notas de lançamento a seguir incluem informações para a versão EMR 5.11.1 da Amazon. As alterações são relativas à EMR versão 5.8.0 da Amazon.

Data da versão inicial: 22 de janeiro de 2018

Alterações, melhorias e problemas resolvidos

- Atualizou o kernel Amazon Linux do default Amazon Linux AMI para Amazon para resolver vulnerabilidades associadas EMR à execução especulativa (CVE-2017-5715, -2017-5753 e -2017-5754). CVE CVE Para ter mais informações, consulte <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

Versões de componente 5.11.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>aws-sagemaker-spark-sdk</code>	1,0	Amazon SageMaker Spark SDK
<code>emr-ddb</code>	4.5.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
emr-goodies	2.4.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.8.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.20.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.3.2	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-6	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.

Componente	Version (Versão)	Descrição
hadoop-hdfs-datanode	2.7.3-amzn-6	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-6	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-6	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-https-server	2.7.3-amzn-6	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-6	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-6	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-6	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-6	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-6	Serviço para recuperar informações atuais e históricas para YARN aplicativos.

Componente	Version (Versão)	Descrição
hbase-hmaster	1.3.1	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.3.1	Serviço para atender uma ou mais HBase regiões.
hbase-client	1.3.1	HBase cliente de linha de comando.
hbase-rest-server	1.3.1	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.3.1	Serviço que fornece um endpoint Thrift para HBase.
hcatalog-client	2.3.2-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.2-amzn-0	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.2-amzn-0	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.3.2-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.3.2-amzn-0	Cliente do Hive-hbase.

Componente	Version (Versão)	Descrição
hive-metastore-server	2.3.2-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.2-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.0.1	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
livy-server	0.4.0-incubating	RESTinterface para interagir com o Apache Spark
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	0.12.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.0.176	Drivers NVIDIA e toolkit CUDA
oozie-client	4.3.0	O cliente da linha de comando do Oozie.
oozie-server	4.3.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.

Componente	Version (Versão)	Descrição
phoenix-library	4.11.0- -1.3 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.11.0- -1.3 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.187	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.187	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
spark-client	2.2.1	Os clientes da linha de comando do Spark.
spark-history-server	2.2.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.2.1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.2.1	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.

Componente	Version (Versão)	Descrição
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.7.3	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.11.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.11.1

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.

Classificações	Descrição
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.

Classificações	Descrição
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure as zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.

Classificações	Descrição
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.

Classificações	Descrição
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.

Classificações	Descrição
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.

Classificações	Descrição
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.

Classificações	Descrição
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.11.0

Versões 5.11.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.11.0	emr-5.10.1	emr-5.10.0	emr-5.9.1
AWS SDK para Java	1.11.238	1.11.221	1.11.221	1.11.183
Python	2,7, 3,4	2,7, 3,4	2,7, 3,4	Não rastreadas
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-

	emr-5.11.0	emr-5.10.1	emr-5.10.0	emr-5.9.1
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	2.3.1	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

	emr-5.11.0	emr-5.10.1	emr-5.10.0	emr-5.9.1
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	0.7.2
ZooKeeper	3.4.10	3.4.10	3.4.10	3.4.10

Notas da versão 5.11.0

As notas de lançamento a seguir incluem informações para a EMR versão 5.11.0 da Amazon. As alterações são referentes à versão 5.10.0.

Atualizações

- Hive 2.3.2
- Spark 2.2.1
- SDK para Java 1.11.238

Novos atributos

- Spark
 - Adicionada a configuração `spark.decommissioning.timeout.threshold`, que aprimora o comportamento de desativação do Spark ao serem usadas instâncias spot. Para ter mais informações, consulte [Configurar o comportamento de desativação de nós](#).
 - [Foi adicionado o `aws-sagemaker-spark-sdk` componente ao Spark, que instala o Amazon SageMaker Spark e as dependências associadas para a integração do Spark com a Amazon SageMaker](#) Você pode usar o Amazon SageMaker Spark para criar pipelines de aprendizado de máquina (ML) do Spark usando os estágios da Amazon SageMaker Para obter mais informações, consulte o [SageMaker Spark Readme](#) sobre GitHub e como [usar o Apache Spark com a Amazon SageMaker no Amazon Developer Guide](#). SageMaker

Problemas conhecidos

- MXNet não inclui bibliotecas OpenCV.
- Por padrão, o Hive 2.3.2 define `hive.compute.query.using.stats=true`. Isso faz com que as consultas obtenham dados de estatísticas existentes em vez de diretamente dos dados, o que pode gerar confusão. Por exemplo, se você tiver uma tabela com `hive.compute.query.using.stats=true` e fizer upload de novos arquivos para a tabela LOCATION, a execução de uma consulta `SELECT COUNT(*)` na tabela retornará a contagem das estatísticas, e não selecionará as linhas adicionadas.

Como alternativa, use o comando `ANALYZE TABLE` para reunir novas estatísticas ou defina `hive.compute.query.using.stats=false`. Para obter mais informações, consulte [Estatísticas no Hive](#) na documentação do Apache Hive.

Versões de componente 5.11.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion-amzn-EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
aws-sagemaker-spark-sdk	1,0	Amazon SageMaker Spark SDK
emr-ddb	4.5.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
emr-goodies	2.4.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.8.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.20.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.3.2	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-6	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.

Componente	Version (Versão)	Descrição
hadoop-hdfs-datanode	2.7.3-amzn-6	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-6	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-6	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-https-server	2.7.3-amzn-6	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-6	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-6	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-6	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-6	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-6	Serviço para recuperar informações atuais e históricas para YARN aplicativos.

Componente	Version (Versão)	Descrição
hbase-hmaster	1.3.1	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.3.1	Serviço para atender uma ou mais HBase regiões.
hbase-client	1.3.1	HBase cliente de linha de comando.
hbase-rest-server	1.3.1	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.3.1	Serviço que fornece um endpoint Thrift para HBase.
hcatalog-client	2.3.2-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.2-amzn-0	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.2-amzn-0	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.3.2-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.3.2-amzn-0	Cliente do Hive-hbase.

Componente	Version (Versão)	Descrição
hive-metastore-server	2.3.2-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.2-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.0.1	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
livy-server	0.4.0-incubating	RESTinterface para interagir com o Apache Spark
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	0.12.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.0.176	Drivers NVIDIA e toolkit CUDA
oozie-client	4.3.0	O cliente da linha de comando do Oozie.
oozie-server	4.3.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.

Componente	Version (Versão)	Descrição
phoenix-library	4.11.0- -1.3 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.11.0- -1.3 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.187	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.187	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
spark-client	2.2.1	Os clientes da linha de comando do Spark.
spark-history-server	2.2.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.2.1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.2.1	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.

Componente	Version (Versão)	Descrição
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.7.3	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.11.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.11.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.

Classificações	Descrição
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.

Classificações	Descrição
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure as zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.

Classificações	Descrição
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.

Classificações	Descrição
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.

Classificações	Descrição
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.

Classificações	Descrição
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.

Classificações	Descrição
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.10.1

Versões 5.10.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.10.1	emr-5.10.0	emr-5.9.1	emr-5.9.0
AWS SDK para Java	1.11.221	1.11.221	1.11.183	1.11.183
Python	2,7, 3,4	2,7, 3,4	Não rastreadas	Não rastreadas
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-

	emr-5.10.1	emr-5.10.0	emr-5.9.1	emr-5.9.0
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
Presto	0.187	0.187	0.184	0.184
Spark	2.2.0	2.2.0	2.2.0	2.2.0

	emr-5.10.1	emr-5.10.0	emr-5.9.1	emr-5.9.0
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

Notas da versão 5.10.1

Esta é uma versão de patch para adicionar a autenticação AWS Signature versão 4 para solicitações ao Amazon S3. Todos os aplicativos e componentes são iguais aos da EMR versão anterior da Amazon.

Important

Nesta versão de lançamento, a Amazon EMR usa o AWS Signature versão 4 exclusivamente para autenticar solicitações para o Amazon S3. Para obter mais informações, consulte [Novidades](#).

Versões de componente 5.10.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`.

O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado myapp-component com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como 2.2-amzn-2.

Componente	Version (Versão)	Descrição
emr-ddb	4.5.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.4.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.7.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.20.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.3.2	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.

Componente	Version (Versão)	Descrição
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-5	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-5	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-5	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-5	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-https-server	2.7.3-amzn-5	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-5	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-5	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-5	YARNserviço para gerenciar contêineres em um nó individual.

Componente	Version (Versão)	Descrição
hadoop-yarn-resourcemanager	2.7.3-amzn-5	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-5	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.3.1	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.3.1	Serviço para atender uma ou mais HBase regiões.
hbase-client	1.3.1	HBasecliente de linha de comando.
hbase-rest-server	1.3.1	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	1.3.1	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	2.3.1-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.1-amzn-0	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.

Componente	Version (Versão)	Descrição
hcatalog-webhcat-server	2.3.1-amzn-0	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.3.1-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.3.1-amzn-0	Cliente do Hive-hbase.
hive-metastore-server	2.3.1-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.1-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.0.1	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
livy-server	0.4.0-incubating	REST interface para interagir com o Apache Spark
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	0.12.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.0.176	Drivers NVIDIA e toolkit CUDA

Componente	Version (Versão)	Descrição
oozie-client	4.3.0	O cliente da linha de comando do Oozie.
oozie-server	4.3.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.11.0- -1.3 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.11.0- -1.3 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.187	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.187	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
spark-client	2.2.0	Os clientes da linha de comando do Spark.
spark-history-server	2.2.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.2.0	Mecanismo de execução na memória para YARN.

Componente	Version (Versão)	Descrição
spark-yarn-slave	2.2.0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.7.3	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.10.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.10.1

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.

Classificações	Descrição
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure as zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.

Classificações	Descrição
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.

Classificações	Descrição
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.

Classificações	Descrição
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.

Classificações	Descrição
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.

Classificações	Descrição
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.10.0

Versões 5.10.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [MXNet](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.10.0	emr-5.9.1	emr-5.9.0	emr-5.8.3
AWS SDK para Java	1.11.221	1.11.183	1.11.183	1.11.160
Python	2,7, 3,4	Não rastreadas	Não rastreadas	Não rastreadas

	emr-5.10.0	emr-5.9.1	emr-5.9.0	emr-5.8.3
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	-	-	-	-
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

	emr-5.10.0	emr-5.9.1	emr-5.9.0	emr-5.8.3
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

Notas da versão 5.10.0

As notas de lançamento a seguir incluem informações para a EMR versão 5.10.0 da Amazon. As alterações são relativas à EMR versão 5.9.0 da Amazon.

Atualizações

- AWS SDK for Java 1.11.21
- Hive 2.3.1
- Presto 0.187

Novos atributos

- Adicionado o suporte para autenticação do Kerberos. Para obter mais informações, consulte [Usar a autenticação Kerberos](#) no Amazon Management Guide EMR
- Foi adicionado suporte para IAM funções de EMRFS. Para obter mais informações, consulte [Configurar IAM funções para EMRFS solicitações ao Amazon S3](#) no Amazon Management Guide EMR

- Foi adicionado suporte para GPU tipos de instância P2 e P3 baseados. Para obter mais informações, consulte Instâncias [EC2P2 da Amazon](#) e [Instâncias EC2 P3 da Amazon](#). NVIDIAo driver 384.81 e o CUDA driver 9.0.176 são instalados nesses tipos de instância por padrão.
- O suporte adicionado para [Apache MXNet](#).

Alterações, melhorias e problemas resolvidos

- Presto
 - Foi adicionado suporte para usar o AWS Glue Data Catalog como metastore padrão do Hive. Para obter mais informações, consulte [Usando o Presto com o AWS Glue Data Catalog](#).
 - Suporte adicionado para [funções geoespaciais](#).
 - Suporte de [vazamento para disco](#) adicionado para uniões.
 - Suporte adicionado para o [Conector Redshift](#).
- Spark
 - Backported [SPARK-20640](#), que torna o tempo limite do rpc e as novas tentativas de valores de registro aleatório configuráveis, usando e propriedades. `spark.shuffle.registration.timeout` `spark.shuffle.registration.maxAttempts`
 - Backported [SPARK-21549](#), que corrige um erro que ocorre ao gravar de forma personalizada OutputFormat em locais diferentes. HDFS
- Backport [Hadoop-13270](#)
- As bibliotecas Numpy, Scipy e Matplotlib foram removidas da Amazon base. EMR AMI Se forem necessárias para o aplicativo, essas bibliotecas estarão disponíveis no repositório do aplicativo. Portanto, você pode usar uma ação de bootstrap para instalá-las em todos os nós usando `yum install`.
- A EMR base da Amazon AMI não tem mais RPM pacotes de aplicativos incluídos, então os RPM pacotes não estão mais presentes nos nós do cluster. A Custom AMIs e a EMR base da Amazon AMI agora fazem referência ao repositório de RPM pacotes no Amazon S3.
- Devido à introdução do faturamento por segundo na AmazonEC2, o comportamento padrão de redução de escala agora é Encerrar na conclusão da tarefa, em vez de Encerrar na hora da instância. Para obter mais informações, consulte [Configure Cluster Scale-Down \(Configurar escalabilidade vertical do cluster\)](#).

Problemas conhecidos

- MXNet não inclui bibliotecas OpenCV.
- Por padrão, o Hive 2.3.1 define `hive.compute.query.using.stats=true`. Isso faz com que as consultas obtenham dados de estatísticas existentes em vez de diretamente dos dados, o que pode gerar confusão. Por exemplo, se você tiver uma tabela com `hive.compute.query.using.stats=true` e fizer upload de novos arquivos para a tabela LOCATION, a execução de uma consulta `SELECT COUNT(*)` na tabela retornará a contagem das estatísticas, e não selecionará as linhas adicionadas.

Como alternativa, use o comando `ANALYZE TABLE` para reunir novas estatísticas ou defina `hive.compute.query.using.stats=false`. Para obter mais informações, consulte [Estatísticas no Hive](#) na documentação do Apache Hive.

Versões de componente 5.10.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion-amzn-EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
emr-ddb	4.5.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.4.0	Bibliotecas convenientes para o ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
emr-kinesis	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.7.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.20.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.3.2	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-5	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-5	HDFSserviço em nível de nó para armazenar blocos.

Componente	Version (Versão)	Descrição
hadoop-hdfs-library	2.7.3-amzn-5	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-5	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httfs-server	2.7.3-amzn-5	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-5	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-5	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-5	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-5	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-5	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.3.1	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.

Componente	Version (Versão)	Descrição
hbase-region-server	1.3.1	Serviço para atender uma ou mais HBase regiões.
hbase-client	1.3.1	HBase cliente de linha de comando.
hbase-rest-server	1.3.1	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.3.1	Serviço que fornece um endpoint Thrift para HBase
hcatalog-client	2.3.1-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.1-amzn-0	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.1-amzn-0	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.3.1-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.3.1-amzn-0	Cliente do Hive-hbase.
hive-metastore-server	2.3.1-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL

Componente	Version (Versão)	Descrição
hive-server2	2.3.1-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.0.1	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
livy-server	0.4.0-incubating	RESTinterface para interagir com o Apache Spark
mahout-client	0.13.0	A biblioteca de Machine Learning.
mxnet	0.12.0	Uma biblioteca flexível, escalável e eficiente para aprendizado aprofundado.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
nvidia-cuda	9.0.176	Drivers NVIDIA e toolkit CUDA
oozie-client	4.3.0	O cliente da linha de comando do Oozie.
oozie-server	4.3.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.11.0- -1.3 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.11.0- -1.3 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API

Componente	Version (Versão)	Descrição
presto-coordinator	0.187	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.187	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
spark-client	2.2.0	Os clientes da linha de comando do Spark.
spark-history-server	2.2.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.2.0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.2.0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.

Componente	Version (Versão)	Descrição
zeppelin-server	0.7.3	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.10.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.10.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.

Classificações	Descrição
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure as zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.

Classificações	Descrição
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.

Classificações	Descrição
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.

Classificações	Descrição
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.

Classificações	Descrição
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.

Classificações	Descrição
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.9.1

Versões 5.9.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.9.1	emr-5.9.0	emr-5.8.3	emr-5.8.2
AWS SDK para Java	1.11.183	1.11.183	1.11.160	1.11.160
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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	-	-	-	-

	emr-5.9.1	emr-5.9.0	emr-5.8.3	emr-5.8.2
Hue	4.0.1	4.0.1	3.12.0	3.12.0
Iceberg	-	-	-	-
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

Notas da versão 5.9.1

Esta é uma versão de patch para adicionar a autenticação AWS Signature versão 4 para solicitações ao Amazon S3. Todos os aplicativos e componentes são iguais aos da EMR versão anterior da Amazon.

Important

Nesta versão de lançamento, a Amazon EMR usa o AWS Signature versão 4 exclusivamente para autenticar solicitações para o Amazon S3. Para obter mais informações, consulte [Novidades](#).

Versões de componente 5.9.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>emr-ddb</code>	4.4.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.4.0	Bibliotecas convenientes para o ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
emr-kinesis	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.7.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.19.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.3.2	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-4	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-4	HDFSserviço em nível de nó para armazenar blocos.

Componente	Version (Versão)	Descrição
hadoop-hdfs-library	2.7.3-amzn-4	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-4	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httfs-server	2.7.3-amzn-4	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-4	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-4	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-4	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-4	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-4	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.3.1	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.

Componente	Version (Versão)	Descrição
hbase-region-server	1.3.1	Serviço para atender uma ou mais HBase regiões.
hbase-client	1.3.1	HBase cliente de linha de comando.
hbase-rest-server	1.3.1	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.3.1	Serviço que fornece um endpoint Thrift para HBase
hcatalog-client	2.3.0-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.0-amzn-0	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.0-amzn-0	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.3.0-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.3.0-amzn-0	Cliente do Hive-hbase.
hive-metastore-server	2.3.0-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL

Componente	Version (Versão)	Descrição
hive-server2	2.3.0-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.0.1	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
livy-server	0.4.0-incubating	RESTinterface para interagir com o Apache Spark
mahout-client	0.13.0	A biblioteca de Machine Learning.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
oozie-client	4.3.0	O cliente da linha de comando do Oozie.
oozie-server	4.3.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.11.0- -1.3 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.11.0- -1.3 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.184	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.

Componente	Version (Versão)	Descrição
presto-worker	0.184	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
spark-client	2.2.0	Os clientes da linha de comando do Spark.
spark-history-server	2.2.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.2.0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.2.0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.
zeppelin-server	0.7.2	O notebook baseado na web que permite a análise de dados interativa.

Componente	Version (Versão)	Descrição
zookeeper-server	3.4.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.9.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.9.1

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink <code>log4j-yarn-session.properties</code> .

Classificações	Descrição
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.

Classificações	Descrição
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue

Classificações	Descrição
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.

Classificações	Descrição
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.

Classificações	Descrição
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.

Classificações	Descrição
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.9.0

Versões 5.9.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Livy](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.9.0	emr-5.8.3	emr-5.8.2	emr-5.8.1
AWS SDK para Java	1.11.183	1.11.160	1.11.160	1.11.160
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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

Notas da versão 5.9.0

As notas de lançamento a seguir incluem informações para a EMR versão 5.9.0 da Amazon. As alterações são relativas à EMR versão 5.8.0 da Amazon.

Data do release: 5 de outubro de 2017

Última atualização de recursos: 12 de outubro de 2017

Atualizações

- AWS SDK for Java versão 1.11.183
- Flink 1.3.2
- Hue 4.0.1
- Pig 0.17.0
- Presto 0.184

Novos atributos

- Adição do suporte ao Livy (versão Livy 0.4.0 - em incubação). Para ter mais informações, consulte [Apache Livy](#).
- Adição de suporte para Hue Notebook para Spark.
- Suporte adicionado para EC2 instâncias Amazon da série i3 (12 de outubro de 2017).

Alterações, melhorias e problemas resolvidos

- Spark
 - Adição de um novo conjunto de recursos que ajudam a garantir que o Spark lide de uma forma mais fácil com o encerramento de nós devido a um redimensionamento manual ou uma solicitação de política de escalabilidade automática. Para ter mais informações, consulte [Configurar o comportamento de desativação de nós](#).
 - SSL é usado em vez de 3 DES para criptografia em trânsito para o serviço de transferência em bloco, o que melhora o desempenho ao usar tipos de EC2 instância da Amazon com AES -NI.
 - Retroportado para [SPARK-21494](#).
- Zeppelin

- Portado para trás [ZEPPELIN-2377](#).
- HBase
 - Foi adicionado o patch [HBASE-18533](#), que permite valores adicionais para HBase BucketCache configuração usando a `hbase-site` classificação de configuração.
- Hue
 - Foi adicionado suporte ao AWS Glue Data Catalog para o editor de consultas Hive no Hue.
 - Por padrão, os superusuários no Hue podem acessar todos os arquivos que as EMR IAM funções da Amazon têm permissão para acessar. Os usuários recém-criados não têm automaticamente permissões para acessar o navegador de arquivos Amazon S3 e devem ter as permissões `filebrowser.s3_access` ativadas para o grupo deles.
- Resolveu um problema que fazia com que JSON os dados subjacentes criados usando o AWS Glue Data Catalog ficassem inacessíveis.

Problemas conhecidos

- A inicialização do cluster falha quando todos os aplicativos são instalados e o tamanho padrão do volume EBS raiz da Amazon não é alterado. Como solução alternativa, use o `aws emr create-cluster` comando do AWS CLI e especifique um `--ebs-root-volume-size` parâmetro maior.
- Por padrão, o Hive 2.3.0 define `hive.compute.query.using.stats=true`. Isso faz com que as consultas obtenham dados de estatísticas existentes em vez de diretamente dos dados, o que pode gerar confusão. Por exemplo, se você tiver uma tabela com `hive.compute.query.using.stats=true` e fizer upload de novos arquivos para a tabela `LOCATION`, a execução de uma consulta `SELECT COUNT(*)` na tabela retornará a contagem das estatísticas, e não selecionará as linhas adicionadas.

Como alternativa, use o comando `ANALYZE TABLE` para reunir novas estatísticas ou defina `hive.compute.query.using.stats=false`. Para obter mais informações, consulte [Estatísticas no Hive](#) na documentação do Apache Hive.

Versões de componente 5.9.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são

a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado myapp-component com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como 2.2-amzn-2.

Componente	Version (Versão)	Descrição
emr-ddb	4.4.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.4.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.7.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.19.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.3.2	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.

Componente	Version (Versão)	Descrição
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-4	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-4	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-4	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-4	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httpfs-server	2.7.3-amzn-4	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-4	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-4	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.

Componente	Version (Versão)	Descrição
hadoop-yarn-nodemanager	2.7.3-amzn-4	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-4	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-4	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.3.1	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.3.1	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.3.1	HBasecliente de linha de comando.
hbase-rest-server	1.3.1	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	1.3.1	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	2.3.0-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.

Componente	Version (Versão)	Descrição
hcatalog-server	2.3.0-amzn-0	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.0-amzn-0	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	2.3.0-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.3.0-amzn-0	Cliente do Hive-hbase.
hive-metastore-server	2.3.0-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.0-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	4.0.1	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
livy-server	0.4.0-incubating	RESTinterface para interagir com o Apache Spark
mahout-client	0.13.0	A biblioteca de Machine Learning.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.

Componente	Version (Versão)	Descrição
oozie-client	4.3.0	O cliente da linha de comando do Oozie.
oozie-server	4.3.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.11.0- -1.3 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.11.0- -1.3 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.184	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.184	O serviço que executa partes de uma consulta.
pig-client	0.17.0	O cliente da linha de comando do Pig.
spark-client	2.2.0	Os clientes da linha de comando do Spark.
spark-history-server	2.2.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.2.0	Mecanismo de execução na memória para YARN.

Componente	Version (Versão)	Descrição
spark-yarn-slave	2.2.0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.7.2	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.9.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.9.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.

Classificações	Descrição
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.

Classificações	Descrição
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
livy-conf	Alterar os valores no arquivo livy.conf do Livy.
livy-env	Alterar os valores no ambiente do Livy.

Classificações	Descrição
livy-log4j	Alterar as configurações de log4j.properties no Livy.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.

Classificações	Descrição
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.

Classificações	Descrição
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.

Classificações	Descrição
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.8.3

Versões 5.8.3 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.8.3	emr-5.8.2	emr-5.8.1	emr-5.8.0
AWS SDK para Java	1.11.160	1.11.160	1.11.160	1.11.160
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
Scala	2.11.8	2.11.8	2.11.8	2.11.8

	emr-5.8.3	emr-5.8.2	emr-5.8.1	emr-5.8.0
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	-	-	-	-
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

	emr-5.8.3	emr-5.8.2	emr-5.8.1	emr-5.8.0
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

Notas da versão 5.8.3

Esta é uma versão de patch para adicionar a autenticação AWS Signature versão 4 para solicitações ao Amazon S3. Todos os aplicativos e componentes são iguais aos da EMR versão anterior da Amazon.

Important

Nesta versão de lançamento, a Amazon EMR usa o AWS Signature versão 4 exclusivamente para autenticar solicitações para o Amazon S3. Para obter mais informações, consulte [Novidades](#).

Versões de componente 5.8.3

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são

a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado myapp-component com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como 2.2-amzn-2.

Componente	Version (Versão)	Descrição
emr-ddb	4.4.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.4.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.6.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.18.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.3.1	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.

Componente	Version (Versão)	Descrição
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-3	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-3	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-3	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-3	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httpfs-server	2.7.3-amzn-3	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-3	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-3	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.

Componente	Version (Versão)	Descrição
hadoop-yarn-nodemanager	2.7.3-amzn-3	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-3	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-3	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.3.1	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.3.1	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.3.1	HBasecliente de linha de comando.
hbase-rest-server	1.3.1	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	1.3.1	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	2.3.0-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.

Componente	Version (Versão)	Descrição
hcatalog-server	2.3.0-amzn-0	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.0-amzn-0	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	2.3.0-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.3.0-amzn-0	Cliente do Hive-hbase.
hive-metastore-server	2.3.0-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.0-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.12.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.13.0	A biblioteca de Machine Learning.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
oozie-client	4.3.0	O cliente da linha de comando do Oozie.

Componente	Version (Versão)	Descrição
oozie-server	4.3.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.11.0- -1.3 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.11.0- -1.3 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.170	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.170	O serviço que executa partes de uma consulta.
pig-client	0.16.0-amzn-1	O cliente da linha de comando do Pig.
spark-client	2.2.0	Os clientes da linha de comando do Spark.
spark-history-server	2.2.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.2.0	Mecanismo de execução na memória para YARN.

Componente	Version (Versão)	Descrição
spark-yarn-slave	2.2.0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.7.2	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.8.3

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.8.3

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.

Classificações	Descrição
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.

Classificações	Descrição
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.

Classificações	Descrição
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.

Classificações	Descrição
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.

Classificações	Descrição
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.

Classificações	Descrição
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.8.2

Versões 5.8.2 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.8.2	emr-5.8.1	emr-5.8.0	emr-5.7.1
AWS SDK para Java	1.11.160	1.11.160	1.11.160	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-

	emr-5.8.2	emr-5.8.1	emr-5.8.0	emr-5.7.1
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	-	-	-	-
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

	emr-5.8.2	emr-5.8.1	emr-5.8.0	emr-5.7.1
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

Notas da versão 5.8.2

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.8.2 da Amazon. As alterações são referentes à versão 5.8.1.

Data da versão inicial: 29 de março de 2018

Alterações, melhorias e problemas resolvidos

- Atualizou o kernel Amazon Linux do default Amazon Linux AMI para Amazon EMR para resolver possíveis vulnerabilidades.

Versões de componente 5.8.2

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
emr-ddb	4.4.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.4.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.6.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.18.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.3.1	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas

Componente	Version (Versão)	Descrição
		coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-3	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-3	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-3	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-3	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httpfs-server	2.7.3-amzn-3	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-3	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-3	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-3	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-3	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.

Componente	Version (Versão)	Descrição
hadoop-yarn-timeline-server	2.7.3-amzn-3	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.3.1	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.3.1	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.3.1	HBase cliente de linha de comando.
hbase-rest-server	1.3.1	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.3.1	Serviço que fornece um endpoint Thrift para HBase
hcatalog-client	2.3.0-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.0-amzn-0	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.0-amzn-0	HTTP endpoint fornecendo uma REST interface para HCatalog.

Componente	Version (Versão)	Descrição
hive-client	2.3.0-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.3.0-amzn-0	Cliente do Hive-hbase.
hive-metastore-server	2.3.0-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.0-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.12.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.13.0	A biblioteca de Machine Learning.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
oozie-client	4.3.0	O cliente da linha de comando do Oozie.
oozie-server	4.3.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.11.0- -1.3 HBase	As bibliotecas do phoenix para servidor e cliente

Componente	Version (Versão)	Descrição
phoenix-query-server	4.11.0- -1.3 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.170	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.170	O serviço que executa partes de uma consulta.
pig-client	0.16.0-amzn-1	O cliente da linha de comando do Pig.
spark-client	2.2.0	Os clientes da linha de comando do Spark.
spark-history-server	2.2.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.2.0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.2.0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.

Componente	Version (Versão)	Descrição
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.7.2	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.8.2

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.8.2

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .

Classificações	Descrição
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.

Classificações	Descrição
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.

Classificações	Descrição
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.

Classificações	Descrição
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.

Classificações	Descrição
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.

Classificações	Descrição
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.8.1

Versões 5.8.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.8.1	emr-5.8.0	emr-5.7.1	emr-5.7.0
AWS SDK para Java	1.11.160	1.11.160	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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	-	-	-	-

	emr-5.8.1	emr-5.8.0	emr-5.7.1	emr-5.7.0
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	4.11.0	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

Notas da versão 5.8.1

As notas de lançamento a seguir incluem informações para a versão EMR 5.8.1 da Amazon. As alterações são relativas à EMR versão 5.8.0 da Amazon.

Data da versão inicial: 22 de janeiro de 2018

Alterações, melhorias e problemas resolvidos

- Atualizou o kernel Amazon Linux do default Amazon Linux AMI para Amazon para resolver vulnerabilidades associadas EMR à execução especulativa (CVE-2017-5715, -2017-5753 e -2017-5754). CVE CVE Para ter mais informações, consulte <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

Versões de componente 5.8.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
emr-ddb	4.4.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.4.0	Bibliotecas convenientes para o ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
emr-kinesis	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.6.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.18.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.3.1	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-3	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-3	HDFSserviço em nível de nó para armazenar blocos.

Componente	Version (Versão)	Descrição
hadoop-hdfs-library	2.7.3-amzn-3	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-3	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httfs-server	2.7.3-amzn-3	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-3	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-3	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-3	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-3	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-3	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.3.1	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.

Componente	Version (Versão)	Descrição
hbase-region-server	1.3.1	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.3.1	HBase cliente de linha de comando.
hbase-rest-server	1.3.1	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.3.1	Serviço que fornece um endpoint Thrift para HBase
hcatalog-client	2.3.0-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.0-amzn-0	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.0-amzn-0	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.3.0-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.3.0-amzn-0	Cliente do Hive-hbase.
hive-metastore-server	2.3.0-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL

Componente	Version (Versão)	Descrição
hive-server2	2.3.0-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.12.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.13.0	A biblioteca de Machine Learning.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
oozie-client	4.3.0	O cliente da linha de comando do Oozie.
oozie-server	4.3.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.11.0- -1.3 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.11.0- -1.3 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.170	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.170	O serviço que executa partes de uma consulta.

Componente	Version (Versão)	Descrição
pig-client	0.16.0-amzn-1	O cliente da linha de comando do Pig.
spark-client	2.2.0	Os clientes da linha de comando do Spark.
spark-history-server	2.2.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.2.0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.2.0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.
zeppelin-server	0.7.2	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.

Componente	Version (Versão)	Descrição
zookeeper-client	3.4.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.8.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.8.1

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Alterar as configurações de <code>log4j-cli.properties</code> no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Hadoop.

Classificações	Descrição
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.

Classificações	Descrição
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.

Classificações	Descrição
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.

Classificações	Descrição
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.

Classificações	Descrição
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.

Classificações	Descrição
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.8.0

Versões 5.8.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.8.0	emr-5.7.1	emr-5.7.0	emr-5.6.1
AWS SDK para Java	1.11.160	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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

Notas da versão 5.8.0

As notas de lançamento a seguir incluem informações para a EMR versão 5.8.0 da Amazon. As alterações são relativas à EMR versão 5.7.0 da Amazon.

Data da versão inicial: 10 de agosto de 2017

Última atualização de recurso: 25 de setembro de 2017

Atualizações

- AWS SDK1.11.160
- Flink 1.3.1
- Hive 2.3.0. Para obter mais informações, consulte [Notas de release](#) no site do Apache Hive.

- Spark 2.2.0. Para obter mais informações, consulte [Notas de release](#) no site do Apache Spark.

Novos atributos

- Adição de suporte para visualização do histórico de aplicativos (25 de setembro de 2017). Para obter mais informações, consulte [Visualização do histórico do aplicativo](#) no Amazon EMR Management Guide.

Alterações, melhorias e problemas resolvidos

- Integração com o AWS Glue Data Catalog
 - Foi adicionada a capacidade do Hive e do Spark de usar SQL o AWS Glue Data Catalog como armazenamento de metadados do Hive. Para obter mais informações, consulte [Usando o AWS Glue Data Catalog como metastore para o Hive](#) e [Use o AWS Glue Data Catalog como metastore para o Spark SQL](#).
- Foi adicionado o histórico do aplicativo aos detalhes do cluster, o que permite visualizar dados históricos dos YARN aplicativos e detalhes adicionais dos aplicativos Spark. Para obter mais informações, consulte [Exibir histórico do aplicativo](#) no Guia EMR de gerenciamento da Amazon.
- Oozie
 - Porta traseira [OOZIE-2748](#).
- Hue
 - [Porta traseira - 5859 HUE](#)
- HBase
 - Patch adicionado para expor a hora de início do servidor HBase mestre por meio de Java Management Extensions (JMX) usando `getMasterInitializedTime`.
 - Adicionado um patch que melhora a hora de início do cluster.

Problemas conhecidos

- A inicialização do cluster falha quando todos os aplicativos são instalados e o tamanho padrão do volume EBS raiz da Amazon não é alterado. Como solução alternativa, use o `aws emr create-cluster` comando do AWS CLI e especifique um `--ebs-root-volume-size` parâmetro maior.
- Por padrão, o Hive 2.3.0 define `hive.compute.query.using.stats=true`. Isso faz com que as consultas obtenham dados de estatísticas existentes em vez de diretamente

dos dados, o que pode gerar confusão. Por exemplo, se você tiver uma tabela com `hive.compute.query.using.stats=true` e fizer upload de novos arquivos para a tabela `LOCATION`, a execução de uma consulta `SELECT COUNT(*)` na tabela retornará a contagem das estatísticas, e não selecionará as linhas adicionadas.

Como alternativa, use o comando `ANALYZE TABLE` para reunir novas estatísticas ou defina `hive.compute.query.using.stats=false`. Para obter mais informações, consulte [Estatísticas no Hive](#) na documentação do Apache Hive.

- Spark: ao usar o Spark, há um problema de vazamento no manipulador de arquivos com o daemon `appusher`, o que pode ocorrer para um trabalho de execução prolongada do Spark depois de várias horas ou dias. Para corrigir o problema, conecte-se ao nó primário e digite `sudo /etc/init.d/appusher stop`. Isso interrompe o daemon `appusher`, que a Amazon EMR reiniciará automaticamente.
- Application history
 - Os dados históricos dos executores inativos do Spark não está disponível.
 - O histórico do aplicativo não está disponível para clusters que usam uma configuração de segurança para habilitar a criptografia em andamento.

Versões de componente 5.8.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
emr-ddb	4.4.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.4.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.4.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.6.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.18.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.3.1	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas

Componente	Version (Versão)	Descrição
		coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-3	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-3	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-3	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-3	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httpfs-server	2.7.3-amzn-3	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-3	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-3	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-3	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-3	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.

Componente	Version (Versão)	Descrição
hadoop-yarn-timeline-server	2.7.3-amzn-3	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.3.1	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.3.1	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.3.1	HBase cliente de linha de comando.
hbase-rest-server	1.3.1	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.3.1	Serviço que fornece um endpoint Thrift para HBase
hcatalog-client	2.3.0-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.3.0-amzn-0	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.3.0-amzn-0	HTTP endpoint fornecendo uma REST interface para HCatalog.

Componente	Version (Versão)	Descrição
hive-client	2.3.0-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.3.0-amzn-0	Cliente do Hive-hbase.
hive-metastore-server	2.3.0-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.3.0-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.12.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.13.0	A biblioteca de Machine Learning.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
oozie-client	4.3.0	O cliente da linha de comando do Oozie.
oozie-server	4.3.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.11.0- -1.3 HBase	As bibliotecas do phoenix para servidor e cliente

Componente	Version (Versão)	Descrição
phoenix-query-server	4.11.0- -1.3 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.170	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.170	O serviço que executa partes de uma consulta.
pig-client	0.16.0-amzn-1	O cliente da linha de comando do Pig.
spark-client	2.2.0	Os clientes da linha de comando do Spark.
spark-history-server	2.2.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.2.0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.2.0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.

Componente	Version (Versão)	Descrição
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.7.2	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.8.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.8.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .

Classificações	Descrição
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.

Classificações	Descrição
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.

Classificações	Descrição
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.

Classificações	Descrição
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.

Classificações	Descrição
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.

Classificações	Descrição
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.7.1

Versões 5.7.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.7.1	emr-5.7.0	emr-5.6.1	emr-5.6.0
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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	-	-	-	-

	emr-5.7.1	emr-5.7.0	emr-5.6.1	emr-5.6.0
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
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

Notas da versão 5.7.1

Esta é uma versão de patch para adicionar a autenticação AWS Signature versão 4 para solicitações ao Amazon S3. Todos os aplicativos e componentes são iguais aos da EMR versão anterior da Amazon.

Important

Nesta versão de lançamento, a Amazon EMR usa o AWS Signature versão 4 exclusivamente para autenticar solicitações para o Amazon S3. Para obter mais informações, consulte [Novidades](#).

Versões de componente 5.7.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>emr-ddb</code>	4.3.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.3.0	Bibliotecas convenientes para o ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
emr-kinesis	3.3.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.5.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.18.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.3.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-2	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFSserviço em nível de nó para armazenar blocos.

Componente	Version (Versão)	Descrição
hadoop-hdfs-library	2.7.3-amzn-2	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httfs-server	2.7.3-amzn-2	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-2	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.3.1	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.

Componente	Version (Versão)	Descrição
hbase-region-server	1.3.1	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.3.1	HBase cliente de linha de comando.
hbase-rest-server	1.3.1	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.3.1	Serviço que fornece um endpoint Thrift para HBase
hcatalog-client	2.1.1-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.1.1-amzn-0	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.1.1-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.1.1-amzn-0	Cliente do Hive-hbase.
hive-metastore-server	2.1.1-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL

Componente	Version (Versão)	Descrição
hive-server2	2.1.1-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.12.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.13.0	A biblioteca de Machine Learning.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
oozie-client	4.3.0	O cliente da linha de comando do Oozie.
oozie-server	4.3.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.11.0- -1.3 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.11.0- -1.3 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.170	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.170	O serviço que executa partes de uma consulta.

Componente	Version (Versão)	Descrição
pig-client	0.16.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	2.1.1	Os clientes da linha de comando do Spark.
spark-history-server	2.1.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.1.1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.1.1	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.
zeppelin-server	0.7.2	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.

Componente	Version (Versão)	Descrição
zookeeper-client	3.4.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.7.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.7.1

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Alterar as configurações de <code>log4j-cli.properties</code> no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Hadoop.

Classificações	Descrição
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.

Classificações	Descrição
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.

Classificações	Descrição
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.

Classificações	Descrição
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.

Classificações	Descrição
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.

Classificações	Descrição
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.7.0

Versões 5.7.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.7.0	emr-5.6.1	emr-5.6.0	emr-5.5.4
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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	-	-	-	-

	emr-5.7.0	emr-5.6.1	emr-5.6.0	emr-5.5.4
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
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

Notas da versão 5.7.0

As notas de lançamento a seguir incluem informações para a versão EMR 5.7.0 da Amazon. As alterações são relativas à EMR versão 5.6.0 da Amazon.

Data do release: 13 de julho de 2017

Atualizações

- Flink 1.3.0
- Phoenix 4.11.0
- Zeppelin 0.7.2

Novos atributos

- Foi adicionada a capacidade de especificar um customAmazon Linux AMI ao criar um cluster. Para obter mais informações, consulte [Usando um personalizado AMI](#).

Alterações, melhorias e problemas resolvidos

- HBase
 - Capacidade adicional para configurar clusters de HBase réplica de leitura. Consulte [Usando um cluster de réplica de leitura](#).
 - Várias correções de erros e melhorias
- Presto: adicionada a capacidade de configurar `node.properties`.
- YARN- capacidade adicional de configurar `container-log4j.properties`
- Sqoop - backported [SQOOP-2880](#), que introduz um argumento que permite definir o diretório temporário do Sqoop.

Versões de componente 5.7.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
emr-ddb	4.3.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
emr-goodies	2.3.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.3.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.5.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.18.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.3.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-2	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.

Componente	Version (Versão)	Descrição
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-2	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httfs-server	2.7.3-amzn-2	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-2	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Serviço para recuperar informações atuais e históricas para YARN aplicativos.

Componente	Version (Versão)	Descrição
hbase-hmaster	1.3.1	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.3.1	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.3.1	HBase cliente de linha de comando.
hbase-rest-server	1.3.1	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.3.1	Serviço que fornece um endpoint Thrift para HBase.
hcatalog-client	2.1.1-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.1.1-amzn-0	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.1.1-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.1.1-amzn-0	Cliente do Hive-hbase.

Componente	Version (Versão)	Descrição
hive-metastore-server	2.1.1-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.1.1-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.12.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.13.0	A biblioteca de Machine Learning.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
oozie-client	4.3.0	O cliente da linha de comando do Oozie.
oozie-server	4.3.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.11.0- -1.3 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.11.0- -1.3 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API

Componente	Version (Versão)	Descrição
presto-coordinator	0.170	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.170	O serviço que executa partes de uma consulta.
pig-client	0.16.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	2.1.1	Os clientes da linha de comando do Spark.
spark-history-server	2.1.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.1.1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.1.1	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.

Componente	Version (Versão)	Descrição
zeppelin-server	0.7.2	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.7.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.7.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.

Classificações	Descrição
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.

Classificações	Descrição
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.

Classificações	Descrição
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.

Classificações	Descrição
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.

Classificações	Descrição
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.

Classificações	Descrição
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.6.1

Versões 5.6.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)

- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.6.1	emr-5.6.0	emr-5.5.4	emr-5.5.3
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-

	emr-5.6.1	emr-5.6.0	emr-5.5.4	emr-5.5.3
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

Notas da versão 5.6.1

Esta é uma versão de patch para adicionar a autenticação AWS Signature versão 4 para solicitações ao Amazon S3. Todos os aplicativos e componentes são iguais aos da EMR versão anterior da Amazon.

⚠ Important

Nesta versão de lançamento, a Amazon EMR usa o AWS Signature versão 4 exclusivamente para autenticar solicitações para o Amazon S3. Para obter mais informações, consulte [Novidades](#).

Versões de componente 5.6.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>emr-ddb</code>	4.3.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.3.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.3.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.5.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.

Componente	Version (Versão)	Descrição
emrfs	2.17.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.2.1	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-2	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-2	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.

Componente	Version (Versão)	Descrição
hadoop-https-server	2.7.3-amzn-2	HTTP endpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-2	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN serviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resource-manager	2.7.3-amzn-2	YARN serviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.3.1	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.3.1	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.3.1	HBase cliente de linha de comando.

Componente	Version (Versão)	Descrição
hbase-rest-server	1.3.1	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.3.1	Serviço que fornece um endpoint Thrift para HBase
hcatalog-client	2.1.1-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.1.1-amzn-0	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.1.1-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.1.1-amzn-0	Cliente do Hive-hbase.
hive-metastore-server	2.1.1-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.1.1-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.

Componente	Version (Versão)	Descrição
hue-server	3.12.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.13.0	A biblioteca de Machine Learning.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
oozie-client	4.3.0	O cliente da linha de comando do Oozie.
oozie-server	4.3.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.9.0- -1.2 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.9.0- -1.2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.170	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.170	O serviço que executa partes de uma consulta.
pig-client	0.16.0-amzn-0	O cliente da linha de comando do Pig.

Componente	Version (Versão)	Descrição
spark-client	2.1.1	Os clientes da linha de comando do Spark.
spark-history-server	2.1.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.1.1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.1.1	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.
zeppelin-server	0.7.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.6.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.6.1

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Alterar as configurações de <code>log4j-cli.properties</code> no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop

Classificações	Descrição
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat

Classificações	Descrição
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.

Classificações	Descrição
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.

Classificações	Descrição
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.

Classificações	Descrição
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.

Classificações	Descrição
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.6.0

Versões 5.6.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.6.0	emr-5.5.4	emr-5.5.3	emr-5.5.2
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
Scala	2.11.8	2.11.8	2.11.8	2.11.8

	emr-5.6.0	emr-5.5.4	emr-5.5.3	emr-5.5.2
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
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

	emr-5.6.0	emr-5.5.4	emr-5.5.3	emr-5.5.2
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

Notas da versão 5.6.0

As notas de lançamento a seguir incluem informações para a versão EMR 5.6.0 da Amazon. As alterações são relativas à versão EMR 5.5.0 da Amazon.

Data do release: 5 de junho de 2017

Atualizações

- Flink 1.2.1
- HBase1.3.1
- Mahout 0.13.0. Essa é a primeira versão do Mahout a oferecer suporte ao Spark 2.x na Amazon EMR versão 5.0 e posterior.
- Spark 2.1.1

Alterações, melhorias e problemas resolvidos

- Presto

- Foi adicionada a capacidade de SSL TLS habilitar/proteger a comunicação entre os nós do Presto ativando a criptografia em trânsito usando uma configuração de segurança. Para obter mais informações, consulte [Criptografia de dados em trânsito](#).
- Enviado para backport o [Presto 7661](#), o que adiciona a opção VERBOSE à instrução EXPLAIN ANALYZE para relatar estatísticas de baixo nível mais detalhadas sobre um plano de consulta.

Versões de componente 5.6.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>emr-ddb</code>	4.3.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.3.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.3.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.5.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.

Componente	Version (Versão)	Descrição
emrfs	2.17.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.2.1	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-2	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-2	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.

Componente	Version (Versão)	Descrição
hadoop-https-server	2.7.3-amzn-2	HTTP endpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-2	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN serviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN serviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.3.1	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.3.1	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.3.1	HBase cliente de linha de comando.

Componente	Version (Versão)	Descrição
hbase-rest-server	1.3.1	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.3.1	Serviço que fornece um endpoint Thrift para HBase
hcatalog-client	2.1.1-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.1.1-amzn-0	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.1.1-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.1.1-amzn-0	Cliente do Hive-hbase.
hive-metastore-server	2.1.1-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.1.1-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.

Componente	Version (Versão)	Descrição
hue-server	3.12.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.13.0	A biblioteca de Machine Learning.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
oozie-client	4.3.0	O cliente da linha de comando do Oozie.
oozie-server	4.3.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.9.0- -1.2 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.9.0- -1.2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.170	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.170	O serviço que executa partes de uma consulta.
pig-client	0.16.0-amzn-0	O cliente da linha de comando do Pig.

Componente	Version (Versão)	Descrição
spark-client	2.1.1	Os clientes da linha de comando do Spark.
spark-history-server	2.1.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.1.1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.1.1	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.
zeppelin-server	0.7.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.6.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.6.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Alterar as configurações de <code>log4j-cli.properties</code> no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop

Classificações	Descrição
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat

Classificações	Descrição
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.

Classificações	Descrição
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.

Classificações	Descrição
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-env	Altere os valores no arquivo presto-env.sh do Presto.
presto-node	Altere os valores no arquivo node.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.

Classificações	Descrição
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.

Classificações	Descrição
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.5.4

Versões 5.5.4 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.5.4	emr-5.5.3	emr-5.5.2	emr-5.5.1
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
Scala	2.11.8	2.11.8	2.11.8	2.11.8

	emr-5.5.4	emr-5.5.3	emr-5.5.2	emr-5.5.1
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
Pig	0.16.0	0.16.0	0.16.0	0.16.0

	emr-5.5.4	emr-5.5.3	emr-5.5.2	emr-5.5.1
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

Notas da versão 5.5.4

Esta é uma versão de patch para adicionar a autenticação AWS Signature versão 4 para solicitações ao Amazon S3. Todos os aplicativos e componentes são iguais aos da EMR versão anterior da Amazon.

Important

Nesta versão de lançamento, a Amazon EMR usa o AWS Signature versão 4 exclusivamente para autenticar solicitações para o Amazon S3. Para obter mais informações, consulte [Novidades](#).

Versões de componente 5.5.4

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são

a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado myapp-component com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como 2.2-amzn-2.

Componente	Version (Versão)	Descrição
emr-ddb	4.3.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.3.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.3.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.5.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.16.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.2.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.

Componente	Version (Versão)	Descrição
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-2	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-2	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httpfs-server	2.7.3-amzn-2	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-2	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.

Componente	Version (Versão)	Descrição
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.3.0	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.3.0	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.3.0	HBasecliente de linha de comando.
hbase-rest-server	1.3.0	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	1.3.0	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	2.1.1-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.

Componente	Version (Versão)	Descrição
hcatalog-server	2.1.1-amzn-0	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	2.1.1-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.1.1-amzn-0	Cliente do Hive-hbase.
hive-metastore-server	2.1.1-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.1.1-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.12.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.12.2	A biblioteca de Machine Learning.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
oozie-client	4.3.0	O cliente da linha de comando do Oozie.

Componente	Version (Versão)	Descrição
oozie-server	4.3.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.9.0- -1.2 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.9.0- -1.2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.170	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.170	O serviço que executa partes de uma consulta.
pig-client	0.16.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	2.1.0	Os clientes da linha de comando do Spark.
spark-history-server	2.1.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.1.0	Mecanismo de execução na memória para YARN.

Componente	Version (Versão)	Descrição
spark-yarn-slave	2.1.0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.7.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.5.4

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.5.4

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.

Classificações	Descrição
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.

Classificações	Descrição
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.

Classificações	Descrição
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.

Classificações	Descrição
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.

Classificações	Descrição
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.5.3

Versões 5.5.3 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.5.3	emr-5.5.2	emr-5.5.1	emr-5.5.0
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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

	emr-5.5.3	emr-5.5.2	emr-5.5.1	emr-5.5.0
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
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-5.5.3	emr-5.5.2	emr-5.5.1	emr-5.5.0
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

Notas da versão 5.5.3

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.5.3 da Amazon. As alterações são referentes à versão 5.5.2.

Data da versão inicial: 29 de agosto de 2018

Alterações, melhorias e problemas resolvidos

- Esta versão aborda uma possível vulnerabilidade de segurança.

Versões de componente 5.5.3

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
emr-ddb	4.3.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.3.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.3.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.5.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.16.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.2.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas

Componente	Version (Versão)	Descrição
		coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-2	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-2	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httpfs-server	2.7.3-amzn-2	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-2	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.

Componente	Version (Versão)	Descrição
hadoop-yarn-timeline-server	2.7.3-amzn-2	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.3.0	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.3.0	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.3.0	HBase cliente de linha de comando.
hbase-rest-server	1.3.0	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.3.0	Serviço que fornece um endpoint Thrift para HBase
hcatalog-client	2.1.1-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.1.1-amzn-0	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP endpoint fornecendo uma REST interface para HCatalog.

Componente	Version (Versão)	Descrição
hive-client	2.1.1-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.1.1-amzn-0	Cliente do Hive-hbase.
hive-metastore-server	2.1.1-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.1.1-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.12.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.12.2	A biblioteca de Machine Learning.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
oozie-client	4.3.0	O cliente da linha de comando do Oozie.
oozie-server	4.3.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.9.0- -1.2 HBase	As bibliotecas do phoenix para servidor e cliente

Componente	Version (Versão)	Descrição
phoenix-query-server	4.9.0- -1.2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.170	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.170	O serviço que executa partes de uma consulta.
pig-client	0.16.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	2.1.0	Os clientes da linha de comando do Spark.
spark-history-server	2.1.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.1.0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.1.0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.

Componente	Version (Versão)	Descrição
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.7.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.5.3

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.5.3

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .

Classificações	Descrição
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.

Classificações	Descrição
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.

Classificações	Descrição
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.

Classificações	Descrição
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.

Classificações	Descrição
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.

Classificações	Descrição
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.5.2

Versões 5.5.2 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)

- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.5.2	emr-5.5.1	emr-5.5.0	emr-5.4.1
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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	3.11.0
Iceberg	-	-	-	-
JupyterEnterpriseGateway	-	-	-	-

	emr-5.5.2	emr-5.5.1	emr-5.5.0	emr-5.4.1
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
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

Notas da versão 5.5.2

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.5.2 da Amazon. As alterações são referentes à versão 5.5.1.

Data da versão inicial: 29 de março de 2018

Alterações, melhorias e problemas resolvidos

- Atualizou o kernel Amazon Linux do default Amazon Linux AMI para Amazon EMR para resolver possíveis vulnerabilidades.

Versões de componente 5.5.2

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>emr-ddb</code>	4.3.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.3.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.3.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.5.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.

Componente	Version (Versão)	Descrição
emrfs	2.16.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.2.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-2	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-2	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.

Componente	Version (Versão)	Descrição
hadoop-https-server	2.7.3-amzn-2	HTTP endpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-2	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARN serviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARN serviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.3.0	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.3.0	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.3.0	HBase cliente de linha de comando.

Componente	Version (Versão)	Descrição
hbase-rest-server	1.3.0	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.3.0	Serviço que fornece um endpoint Thrift para HBase
hcatalog-client	2.1.1-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.1.1-amzn-0	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.1.1-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.1.1-amzn-0	Cliente do Hive-hbase.
hive-metastore-server	2.1.1-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.1.1-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.

Componente	Version (Versão)	Descrição
hue-server	3.12.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.12.2	A biblioteca de Machine Learning.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
oozie-client	4.3.0	O cliente da linha de comando do Oozie.
oozie-server	4.3.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.9.0- -1.2 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.9.0- -1.2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.170	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.170	O serviço que executa partes de uma consulta.
pig-client	0.16.0-amzn-0	O cliente da linha de comando do Pig.

Componente	Version (Versão)	Descrição
spark-client	2.1.0	Os clientes da linha de comando do Spark.
spark-history-server	2.1.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.1.0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.1.0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.
zeppelin-server	0.7.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.5.2

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.5.2

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Alterar as configurações de <code>log4j-cli.properties</code> no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop

Classificações	Descrição
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat

Classificações	Descrição
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.

Classificações	Descrição
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.

Classificações	Descrição
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.

Classificações	Descrição
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.

Classificações	Descrição
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.5.1

Versões 5.5.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.5.1	emr-5.5.0	emr-5.4.1	emr-5.4.0
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-

	emr-5.5.1	emr-5.5.0	emr-5.4.1	emr-5.4.0
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
Sqoop	1.4.6	1.4.6	1.4.6	1.4.6

	emr-5.5.1	emr-5.5.0	emr-5.4.1	emr-5.4.0
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

Notas da versão 5.5.1

As notas de lançamento a seguir incluem informações para a versão EMR 5.5.1 da Amazon. As alterações são relativas à versão EMR 5.5.0 da Amazon.

Data da versão inicial: 22 de janeiro de 2018

Alterações, melhorias e problemas resolvidos

- Atualizou o kernel Amazon Linux do default Amazon Linux AMI para Amazon para resolver vulnerabilidades associadas EMR à execução especulativa (CVE-2017-5715, -2017-5753 e -2017-5754). CVE CVE Para ter mais informações, consulte <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

Versões de componente 5.5.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto

nomeado myapp-component com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como 2.2-amzn-2.

Componente	Version (Versão)	Descrição
emr-ddb	4.3.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.3.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.3.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.5.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.16.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.2.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.

Componente	Version (Versão)	Descrição
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-2	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-2	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-https-server	2.7.3-amzn-2	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-2	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARNserviço para gerenciar contêineres em um nó individual.

Componente	Version (Versão)	Descrição
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.3.0	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.3.0	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.3.0	HBasecliente de linha de comando.
hbase-rest-server	1.3.0	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	1.3.0	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	2.1.1-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.1.1-amzn-0	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.

Componente	Version (Versão)	Descrição
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.1.1-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.1.1-amzn-0	Cliente do Hive-hbase.
hive-metastore-server	2.1.1-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.1.1-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.12.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.12.2	A biblioteca de Machine Learning.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
oozie-client	4.3.0	O cliente da linha de comando do Oozie.
oozie-server	4.3.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.

Componente	Version (Versão)	Descrição
phoenix-library	4.9.0- -1.2 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.9.0- -1.2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.170	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.170	O serviço que executa partes de uma consulta.
pig-client	0.16.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	2.1.0	Os clientes da linha de comando do Spark.
spark-history-server	2.1.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.1.0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.1.0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.

Componente	Version (Versão)	Descrição
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.7.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.5.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.5.1

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.

Classificações	Descrição
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.

Classificações	Descrição
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.

Classificações	Descrição
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.

Classificações	Descrição
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.

Classificações	Descrição
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.

Classificações	Descrição
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.5.0

Versões 5.5.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.5.0	emr-5.4.1	emr-5.4.0	emr-5.3.2
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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	1.1.4
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

	emr-5.5.0	emr-5.4.1	emr-5.4.0	emr-5.3.2
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
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

Notas da versão 5.5.0

As notas de lançamento a seguir incluem informações para a versão EMR 5.5.0 da Amazon. As alterações são relativas à versão EMR 5.4.0 da Amazon.

Data do release: 26 de abril de 2017

Atualizações

- Hue 3.12
- Presto 0.170
- Zeppelin 0.7.1
- ZooKeeper 3.4.10

Alterações, melhorias e problemas resolvidos

- Spark
 - Backported Spark Patch ([SPARK-20115](#)) [Correção DAGScheduler para recalcular todos os blocos de reprodução aleatória perdidos quando o serviço externo de reprodução aleatória não está disponível na versão 2.1.0 do Spark, incluída](#) nesta versão.
- Flink
 - O Flink agora é compilado com o Scala 2.11. Se você usa o Scala API e as bibliotecas, recomendamos usar o Scala 2.11 em seus projetos.
 - Tratado um problema em que os padrões HADOOP_CONF_DIR e YARN_CONF_DIR não estavam definidos corretamente, portanto havia falha no funcionamento de `start-scala-shell.sh`. Também foi adicionada a possibilidade de definir esses valores usando `env.hadoop.conf.dir` e `env.yarn.conf.dir` em `/etc/flink/conf/flink-conf.yaml` ou na classificação de configuração `flink-conf`.
 - Introduziu um novo comando EMR específico, `flink-scala-shell` como invólucro para `start-scala-shell.sh`. Recomendamos o uso desse comando, em vez de `start-scala-shell`. O novo comando simplifica a execução. Por exemplo, `flink-scala-shell -n 2` inicia um shell Scala Flink com um paralelismo de tarefa de 2.
 - Introduziu um novo comando EMR específico, `flink-yarn-session` como invólucro para `yarn-session.sh`. Recomendamos o uso desse comando, em vez de `yarn-session`. O novo comando simplifica a execução. Por exemplo, `flink-yarn-session -d -n 2` inicia

uma sessão de longa execução do Flink em um estado desanexado com dois gerenciadores de tarefas.

- O [httpclient endereçado \(FLINK-6125\) do Commons não está mais sombreado](#) no Flink 1.2.
- Presto
 - Foi adicionado suporte para LDAP autenticação. O uso LDAP com o Presto na Amazon EMR exige que você habilite o HTTPS acesso para o coordenador do Presto (`http-server.https.enabled=trueemconfig.properties`). Para obter detalhes de configuração, consulte [LDAPAutenticação](#) na documentação do Presto.
 - O suporte adicionado para SHOW GRANTS.
- Amazon EMR Base Linux AMI
 - Os EMR lançamentos da Amazon agora são baseados no Amazon Linux 2017.03. Para obter mais informações, consulte as notas de [versão do Amazon Linux AMI 2017.03](#).
 - O Python 2.6 foi removido da imagem Linux EMR básica da Amazon. Python 2.7 e 3.4 estão instalados por padrão. Você pode instalar o Python 2.6 manualmente, se necessário.

Versões de componente 5.5.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-`amzn-EmrVersion`. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>emr-ddb</code>	4.3.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
emr-goodies	2.3.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.3.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.5.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.16.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.2.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-2	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.

Componente	Version (Versão)	Descrição
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-2	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httfs-server	2.7.3-amzn-2	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-2	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Serviço para recuperar informações atuais e históricas para YARN aplicativos.

Componente	Version (Versão)	Descrição
hbase-hmaster	1.3.0	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.3.0	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.3.0	HBase cliente de linha de comando.
hbase-rest-server	1.3.0	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.3.0	Serviço que fornece um endpoint Thrift para HBase.
hcatalog-client	2.1.1-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.1.1-amzn-0	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.1.1-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.1.1-amzn-0	Cliente do Hive-hbase.

Componente	Version (Versão)	Descrição
hive-metastore-server	2.1.1-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.1.1-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.12.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.12.2	A biblioteca de Machine Learning.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
oozie-client	4.3.0	O cliente da linha de comando do Oozie.
oozie-server	4.3.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.9.0- -1.2 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.9.0- -1.2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API

Componente	Version (Versão)	Descrição
presto-coordinator	0.170	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.170	O serviço que executa partes de uma consulta.
pig-client	0.16.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	2.1.0	Os clientes da linha de comando do Spark.
spark-history-server	2.1.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.1.0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.1.0	Bibliotecas do Apache Spark necessárias para escravos YARN.
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.

Componente	Version (Versão)	Descrição
zeppelin-server	0.7.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.10	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.10	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.5.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.5.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.

Classificações	Descrição
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.

Classificações	Descrição
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.

Classificações	Descrição
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.

Classificações	Descrição
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.

Classificações	Descrição
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.

Classificações	Descrição
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.4.1

Versões 5.4.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.4.1	emr-5.4.0	emr-5.3.2	emr-5.3.1
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
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	-	-	-	-

	emr-5.4.1	emr-5.4.0	emr-5.3.2	emr-5.3.1
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
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

Notas da versão 5.4.1

Esta é uma versão de patch para adicionar a autenticação AWS Signature versão 4 para solicitações ao Amazon S3. Todos os aplicativos e componentes são iguais aos da EMR versão anterior da Amazon.

Important

Nesta versão de lançamento, a Amazon EMR usa o AWS Signature versão 4 exclusivamente para autenticar solicitações para o Amazon S3. Para obter mais informações, consulte [Novidades](#).

Versões de componente 5.4.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion-amzn-EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>emr-ddb</code>	4.2.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.3.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.2.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.4.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
<code>emrfs</code>	2.15.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
<code>flink-client</code>	1.2.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.

Componente	Version (Versão)	Descrição
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-1	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httpfs-server	2.7.3-amzn-1	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API

Componente	Version (Versão)	Descrição
hadoop-mapred	2.7.3-amzn-1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.3.0	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.3.0	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.3.0	HBasecliente de linha de comando.
hbase-rest-server	1.3.0	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	1.3.0	Serviço que fornece um endpoint Thrift para. HBase

Componente	Version (Versão)	Descrição
hcatalog-client	2.1.1-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.1.1-amzn-0	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	2.1.1-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.1.1-amzn-0	Cliente do Hive-hbase.
hive-metastore-server	2.1.1-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.1.1-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.11.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.12.2	A biblioteca de Machine Learning.

Componente	Version (Versão)	Descrição
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
oozie-client	4.3.0	O cliente da linha de comando do Oozie.
oozie-server	4.3.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.9.0- -1.2 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.9.0- -1.2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.166	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.166	O serviço que executa partes de uma consulta.
pig-client	0.16.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	2.1.0	Os clientes da linha de comando do Spark.
spark-history-server	2.1.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.

Componente	Version (Versão)	Descrição
spark-on-yarn	2.1.0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.1.0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.
zeppelin-server	0.7.0	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.9	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.9	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.4.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.4.1

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.

Classificações	Descrição
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.

Classificações	Descrição
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.

Classificações	Descrição
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.

Classificações	Descrição
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.

Classificações	Descrição
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.4.0

Versões 5.4.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.4.0	emr-5.3.2	emr-5.3.1	emr-5.3.0
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
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

	emr-5.4.0	emr-5.3.2	emr-5.3.1	emr-5.3.0
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	3.11.0	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
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-5.4.0	emr-5.3.2	emr-5.3.1	emr-5.3.0
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

Notas da versão 5.4.0

As notas de lançamento a seguir incluem informações para a versão EMR 5.4.0 da Amazon. As alterações são relativas à versão EMR 5.3.0 da Amazon.

Data do release: 08 de março de 2017

Atualizações

- Atualizado para Flink 1.2.0
- Atualizado para 1.3.0 HBase
- Atualizado para Phoenix 4.9.0

Note

Se você atualizar de uma versão anterior da Amazon EMR para a EMR versão 5.4.0 ou posterior da Amazon e usar indexação secundária, atualize os índices locais conforme descrito na documentação do [Apache](#) Phoenix. A Amazon EMR remove as configurações necessárias da `hbase-site` classificação, mas os índices precisam ser preenchidos novamente. O sistema oferece suporte a atualizações de índices online e offline. As atualizações online são o padrão, o que significa que os índices são preenchidos novamente durante a inicialização de clientes do Phoenix versão 4.8.0 ou posterior. Para especificar atualizações off-line, defina a `phoenix.client.localIndexUpgrade` configuração como falsa na `phoenix-site` classificação e, em seguida, como o nó primário SSH a ser executado `psql [zookeeper] -1`.

- Atualizado para Presto 0.166
- Atualizado para Zeppelin 0.7.0

Alterações e melhorias

- Adicionado o suporte para instâncias r4. Consulte os [tipos de EC2 instância da Amazon](#).

Versões de componente 5.4.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>emr-ddb</code>	4.2.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.3.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.2.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.4.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
<code>emrfs</code>	2.15.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
flink-client	1.2.0	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-1	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-https-server	2.7.3-amzn-1	HTTPendpoint para HDFS operações.

Componente	Version (Versão)	Descrição
hadoop-kms-server	2.7.3-amzn-1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.3.0	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.3.0	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.3.0	HBasecliente de linha de comando.
hbase-rest-server	1.3.0	Serviço que fornece um RESTful HTTP endpoint paraHBase.

Componente	Version (Versão)	Descrição
hbase-thrift-server	1.3.0	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	2.1.1-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.1.1-amzn-0	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTPendpoint fornecendo o uma REST interface paraHCatalog.
hive-client	2.1.1-amzn-0	O cliente da linha de comando do Hive.
hive-hbase	2.1.1-amzn-0	Cliente do Hive-hbase.
hive-metastore-server	2.1.1-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server2	2.1.1-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.11.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop

Componente	Version (Versão)	Descrição
mahout-client	0.12.2	A biblioteca de Machine Learning.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
oozie-client	4.3.0	O cliente da linha de comando do Oozie.
oozie-server	4.3.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.9.0- -1.2 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.9.0- -1.2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.166	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.166	O serviço que executa partes de uma consulta.
pig-client	0.16.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	2.1.0	Os clientes da linha de comando do Spark.

Componente	Version (Versão)	Descrição
spark-history-server	2.1.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.1.0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.1.0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.
zeppelin-server	0.7.0	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.9	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.9	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.4.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.4.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Alterar as configurações de <code>log4j-cli.properties</code> no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop

Classificações	Descrição
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat

Classificações	Descrição
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.

Classificações	Descrição
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.

Classificações	Descrição
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.

Classificações	Descrição
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.

Classificações	Descrição
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.3.2

Versões 5.3.2 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.3.2	emr-5.3.1	emr-5.3.0	emr-5.2.3
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-

	emr-5.3.2	emr-5.3.1	emr-5.3.0	emr-5.2.3
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	-	-	-	-
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

	emr-5.3.2	emr-5.3.1	emr-5.3.0	emr-5.2.3
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

Notas da versão 5.3.2

Esta é uma versão de patch para adicionar a autenticação AWS Signature versão 4 para solicitações ao Amazon S3. Todos os aplicativos e componentes são iguais aos da EMR versão anterior da Amazon.

Important

Nesta versão de lançamento, a Amazon EMR usa o AWS Signature versão 4 exclusivamente para autenticar solicitações para o Amazon S3. Para obter mais informações, consulte [Novidades](#).

Versões de componente 5.3.2

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto

nomeado myapp-component com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como 2.2-amzn-2.

Componente	Version (Versão)	Descrição
emr-ddb	4.2.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.2.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.2.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.4.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.14.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.1.4	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.

Componente	Version (Versão)	Descrição
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-1	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-https-server	2.7.3-amzn-1	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARNserviço para gerenciar contêineres em um nó individual.

Componente	Version (Versão)	Descrição
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.2.3	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.2.3	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.2.3	HBasecliente de linha de comando.
hbase-rest-server	1.2.3	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	1.2.3	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	2.1.1-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.1.1-amzn-0	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.

Componente	Version (Versão)	Descrição
hcatalog-webhcat-server	2.1.1-amzn-0	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.1.1-amzn-0	O cliente da linha de comando do Hive.
hive-metastore-server	2.1.1-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	2.1.1-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.11.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.12.2	A biblioteca de Machine Learning.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
oozie-client	4.3.0	O cliente da linha de comando do Oozie.
oozie-server	4.3.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.7.0- -1,2 HBase	As bibliotecas do phoenix para servidor e cliente

Componente	Version (Versão)	Descrição
phoenix-query-server	4.7.0- -1,2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.157.1	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.157.1	O serviço que executa partes de uma consulta.
pig-client	0.16.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	2.1.0	Os clientes da linha de comando do Spark.
spark-history-server	2.1.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.1.0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.1.0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.

Componente	Version (Versão)	Descrição
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.6.2	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.9	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.9	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.3.2

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.3.2

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .

Classificações	Descrição
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.

Classificações	Descrição
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.

Classificações	Descrição
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.

Classificações	Descrição
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.

Classificações	Descrição
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.

Classificações	Descrição
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.3.1

Versões 5.3.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)

- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.3.1	emr-5.3.0	emr-5.2.3	emr-5.2.2
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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	-	-	-	-

	emr-5.3.1	emr-5.3.0	emr-5.2.3	emr-5.2.2
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
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

Notas da versão 5.3.1

As notas de lançamento a seguir incluem informações para a versão EMR 5.3.1 da Amazon. As alterações são relativas à versão EMR 5.3.0 da Amazon.

Data da versão: 7 de fevereiro de 2017

Pequenas alterações nos patches do Zeppelin e na atualização do padrão para a Amazon. AMI EMR

Versões de componente 5.3.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>emr-ddb</code>	4.2.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.2.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.2.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.4.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
<code>emrfs</code>	2.14.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
<code>flink-client</code>	1.1.4	Scripts do cliente da linha de comando e aplicativos do Apache Flink.

Componente	Version (Versão)	Descrição
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-1	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httpfs-server	2.7.3-amzn-1	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API

Componente	Version (Versão)	Descrição
hadoop-mapred	2.7.3-amzn-1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.2.3	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.2.3	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.2.3	HBasecliente de linha de comando.
hbase-rest-server	1.2.3	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	1.2.3	Serviço que fornece um endpoint Thrift para. HBase

Componente	Version (Versão)	Descrição
hcatalog-client	2.1.1-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.1.1-amzn-0	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTPendpoint fornecendo o uma REST interface paraHCatalog.
hive-client	2.1.1-amzn-0	O cliente da linha de comando do Hive.
hive-metastore-server	2.1.1-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	2.1.1-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.11.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.12.2	A biblioteca de Machine Learning.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.

Componente	Version (Versão)	Descrição
oozie-client	4.3.0	O cliente da linha de comando do Oozie.
oozie-server	4.3.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.7.0- -1,2 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.7.0- -1,2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.157.1	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.157.1	O serviço que executa partes de uma consulta.
pig-client	0.16.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	2.1.0	Os clientes da linha de comando do Spark.
spark-history-server	2.1.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.1.0	Mecanismo de execução na memória para YARN.

Componente	Version (Versão)	Descrição
spark-yarn-slave	2.1.0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.6.2	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.9	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.9	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.3.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.3.1

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.

Classificações	Descrição
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.

Classificações	Descrição
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.

Classificações	Descrição
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.

Classificações	Descrição
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.

Classificações	Descrição
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.3.0

Versões 5.3.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.3.0	emr-5.2.3	emr-5.2.2	emr-5.2.1
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
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

	emr-5.3.0	emr-5.2.3	emr-5.2.2	emr-5.2.1
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
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-5.3.0	emr-5.2.3	emr-5.2.2	emr-5.2.1
Trino (PrestoSQL)	-	-	-	-
Zeppelin	0.6.2	0.6.2	0.6.2	0.6.2
ZooKeeper	3.4.9	3.4.9	3.4.9	3.4.9

Notas da versão 5.3.0

As notas de lançamento a seguir incluem informações para a versão EMR 5.3.0 da Amazon. As alterações são relativas à versão EMR 5.2.1 da Amazon.

Data do release: 26 de janeiro de 2017

Atualizações

- Atualizado para Hive 2.1.1
- Atualizado para Hue 3.11.0
- Atualizado para Spark 2.1.0
- Atualizado para Oozie 4.3.0
- Atualizado para Flink 1.1.4

Alterações e melhorias

- Adicionado um patch para o Hue que permite usar a configuração `interpreters_shown_on_wheel` para definir o que intérpretes mostram primeiro na roda de seleção do bloco de anotações, independentemente de sua ordem no arquivo `hue.ini`.
- Adicionada a classificação de configuração `hive-parquet-logging`, que pode ser usada para configurar os valores no arquivo `parquet-logging.properties` do Hive.

Versões de componente 5.3.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou

aws. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado myapp-component com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como 2.2-amzn-2.

Componente	Version (Versão)	Descrição
emr-ddb	4.2.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.2.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.2.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.4.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.14.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.1.4	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.

Componente	Version (Versão)	Descrição
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-1	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httpfs-server	2.7.3-amzn-1	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.

Componente	Version (Versão)	Descrição
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.2.3	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.2.3	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.2.3	HBasecliente de linha de comando.
hbase-rest-server	1.2.3	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	1.2.3	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	2.1.1-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.

Componente	Version (Versão)	Descrição
hcatalog-server	2.1.1-amzn-0	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.1.1-amzn-0	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	2.1.1-amzn-0	O cliente da linha de comando do Hive.
hive-metastore-server	2.1.1-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	2.1.1-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.11.0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.12.2	A biblioteca de Machine Learning.
mysql-server	5.5.52	Meu servidor SQL de banco de dados.
oozie-client	4.3.0	O cliente da linha de comando do Oozie.

Componente	Version (Versão)	Descrição
oozie-server	4.3.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.7.0- -1,2 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.7.0- -1,2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.157.1	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.157.1	O serviço que executa partes de uma consulta.
pig-client	0.16.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	2.1.0	Os clientes da linha de comando do Spark.
spark-history-server	2.1.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.1.0	Mecanismo de execução na memória para YARN.

Componente	Version (Versão)	Descrição
spark-yarn-slave	2.1.0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.6.2	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.9	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.9	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.3.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.3.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.

Classificações	Descrição
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-parquet-logging	Alterar os valores no arquivo parquet-logging.properties do Hive.

Classificações	Descrição
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.

Classificações	Descrição
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.

Classificações	Descrição
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.

Classificações	Descrição
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.2.3

Versões 5.2.3 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.2.3	emr-5.2.2	emr-5.2.1	emr-5.2.0
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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-5.2.3	emr-5.2.2	emr-5.2.1	emr-5.2.0
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	0.157.1	0.157.1	0.152.3
Spark	2.0.2	2.0.2	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

	emr-5.2.3	emr-5.2.2	emr-5.2.1	emr-5.2.0
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

Notas da versão 5.2.3

Esta é uma versão de patch para adicionar a autenticação AWS Signature versão 4 para solicitações ao Amazon S3. Todos os aplicativos e componentes são iguais aos da EMR versão anterior da Amazon.

Important

Nesta versão de lançamento, a Amazon EMR usa o AWS Signature versão 4 exclusivamente para autenticar solicitações para o Amazon S3. Para obter mais informações, consulte [Novidades](#).

Versões de componente 5.2.3

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
emr-ddb	4.2.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.2.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.2.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.4.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.13.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.1.3	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas

Componente	Version (Versão)	Descrição
		coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-1	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-https-server	2.7.3-amzn-1	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.

Componente	Version (Versão)	Descrição
hadoop-yarn-timeline-server	2.7.3-amzn-1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.2.3	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.2.3	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.2.3	HBase cliente de linha de comando.
hbase-rest-server	1.2.3	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.2.3	Serviço que fornece um endpoint Thrift para HBase
hcatalog-client	2.1.0-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.1.0-amzn-0	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP endpoint fornecendo uma REST interface para HCatalog.

Componente	Version (Versão)	Descrição
hive-client	2.1.0-amzn-0	O cliente da linha de comando do Hive.
hive-metastore-server	2.1.0-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	2.1.0-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.10.0-amzn-0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.12.2	A biblioteca de Machine Learning.
mysql-server	5.5.52	Meu servidor SQL de banco de dados.
oozie-client	4.2.0	O cliente da linha de comando do Oozie.
oozie-server	4.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.7.0- -1,2 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.7.0- -1,2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API

Componente	Version (Versão)	Descrição
presto-coordinator	0.157.1	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.157.1	O serviço que executa partes de uma consulta.
pig-client	0.16.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	2.0.2	Os clientes da linha de comando do Spark.
spark-history-server	2.0.2	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.0.2	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.0.2	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.23	HTTP Servidor Apache.

Componente	Version (Versão)	Descrição
zeppelin-server	0.6.2	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.9	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.9	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.2.3

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.2.3

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.

Classificações	Descrição
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.

Classificações	Descrição
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue

Classificações	Descrição
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.

Classificações	Descrição
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.

Classificações	Descrição
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.

Classificações	Descrição
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.2.2

Versões 5.2.2 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.2.2	emr-5.2.1	emr-5.2.0	emr-5.1.1
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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
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-5.2.2	emr-5.2.1	emr-5.2.0	emr-5.1.1
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
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

Notas da versão 5.2.2

As notas de lançamento a seguir incluem informações para a versão EMR 5.2.2 da Amazon. As alterações são relativas à versão EMR 5.2.1 da Amazon.

Data do release: 2 de maio de 2017

Problemas conhecidos das versões anteriores que foram resolvidos

- Backported [SPARK-194459](#), que resolve um problema em que a leitura de uma ORC tabela com colunas char/varchar pode falhar.

Versões de componente 5.2.2

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion-amzn-EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>emr-ddb</code>	4.2.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.2.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.2.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.4.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
<code>emrfs</code>	2.13.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
<code>flink-client</code>	1.1.3	Scripts do cliente da linha de comando e aplicativos do Apache Flink.

Componente	Version (Versão)	Descrição
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-1	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httpfs-server	2.7.3-amzn-1	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API

Componente	Version (Versão)	Descrição
hadoop-mapred	2.7.3-amzn-1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.2.3	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.2.3	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.2.3	HBasecliente de linha de comando.
hbase-rest-server	1.2.3	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	1.2.3	Serviço que fornece um endpoint Thrift para. HBase

Componente	Version (Versão)	Descrição
hcatalog-client	2.1.0-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.1.0-amzn-0	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.1.0-amzn-0	HTTPendpoint fornecendo o uma REST interface paraHCatalog.
hive-client	2.1.0-amzn-0	O cliente da linha de comando do Hive.
hive-metastore-server	2.1.0-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	2.1.0-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.10.0-amzn-0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.12.2	A biblioteca de Machine Learning.
mysql-server	5.5.52	Meu servidor SQL de banco de dados.

Componente	Version (Versão)	Descrição
oozie-client	4.2.0	O cliente da linha de comando do Oozie.
oozie-server	4.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.7.0- -1,2 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.7.0- -1,2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.157.1	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.157.1	O serviço que executa partes de uma consulta.
pig-client	0.16.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	2.0.2	Os clientes da linha de comando do Spark.
spark-history-server	2.0.2	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.0.2	Mecanismo de execução na memória para YARN.

Componente	Version (Versão)	Descrição
spark-yarn-slave	2.0.2	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.23	HTTPServidor Apache.
zeppelin-server	0.6.2	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.9	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.9	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.2.2

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.2.2

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.

Classificações	Descrição
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores no HDFS hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.

Classificações	Descrição
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.

Classificações	Descrição
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.

Classificações	Descrição
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.

Classificações	Descrição
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oracoop-site	Altere os valores no arquivo oracoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.2.1

Versões 5.2.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.2.1	emr-5.2.0	emr-5.1.1	emr-5.1.0
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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

Notas da versão 5.2.1

As notas de lançamento a seguir incluem informações para a versão EMR 5.2.1 da Amazon. As alterações são relativas à versão EMR 5.2.0 da Amazon.

Data do release: 29 de dezembro de 2016

Atualizações

- Atualizado para Presto 0.157.1. Para obter mais informações, consulte [Notas de release do Presto](#) na documentação do Presto.
- Zookeeper atualizado para 3.4.9. Para obter mais informações, consulte as [notas de ZooKeeper lançamento](#) na ZooKeeper documentação do Apache.

Alterações e melhorias

- Foi adicionado suporte para o tipo de EC2 instância Amazon m4.16xlarge na Amazon EMR versão 4.8.3 e posterior, excluindo 5.0.0, 5.0.3 e 5.2.0.
- Os EMR lançamentos da Amazon agora são baseados no Amazon Linux 2016.09. Para ter mais informações, consulte <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.
- A localização do Flink e os caminhos de YARN configuração agora são definidos por padrão, pois `/etc/default/flink` você não precisa definir as variáveis de ambiente `FLINK_CONF_DIR` e `HADOOP_CONF_DIR` ao executar os scripts `flink` ou os scripts do `yarn-session.sh` driver, para iniciar trabalhos do Flink.
- Foi adicionado suporte para a FlinkKinesisConsumer aula.

Problemas conhecidos das versões anteriores que foram resolvidos

- Corrigido um problema no Hadoop em que o ReplicationMonitor thread podia ficar preso por um longo tempo devido a uma corrida entre a replicação e a exclusão do mesmo arquivo em um grande cluster.
- Corrigido um problema em que ControlledJob # toString falhava com uma exceção de ponteiro nulo (NPE) quando o status do trabalho não era atualizado com êxito.

Versões de componente 5.2.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>emr-ddb</code>	4.2.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.2.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.2.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.4.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.

Componente	Version (Versão)	Descrição
emrfs	2.13.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.1.3	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-1	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.

Componente	Version (Versão)	Descrição
hadoop-https-server	2.7.3-amzn-1	HTTP endpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARN serviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARN serviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.2.3	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.2.3	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.2.3	HBase cliente de linha de comando.

Componente	Version (Versão)	Descrição
hbase-rest-server	1.2.3	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.2.3	Serviço que fornece um endpoint Thrift para HBase
hcatalog-client	2.1.0-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.1.0-amzn-0	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.1.0-amzn-0	O cliente da linha de comando do Hive.
hive-metastore-server	2.1.0-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	2.1.0-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.10.0-amzn-0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop

Componente	Version (Versão)	Descrição
mahout-client	0.12.2	A biblioteca de Machine Learning.
mysql-server	5.5.52	Meu servidor SQL de banco de dados.
oozie-client	4.2.0	O cliente da linha de comando do Oozie.
oozie-server	4.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.7.0- -1,2 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.7.0- -1,2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.157.1	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.157.1	O serviço que executa partes de uma consulta.
pig-client	0.16.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	2.0.2	Os clientes da linha de comando do Spark.

Componente	Version (Versão)	Descrição
spark-history-server	2.0.2	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.0.2	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.0.2	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.23	HTTP Servidor Apache.
zeppelin-server	0.6.2	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.9	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.9	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.2.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.2.1

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Alterar as configurações de <code>log4j-cli.properties</code> no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop

Classificações	Descrição
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat

Classificações	Descrição
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.

Classificações	Descrição
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.

Classificações	Descrição
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.

Classificações	Descrição
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.2.0

Versões 5.2.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.2.0	emr-5.1.1	emr-5.1.0	emr-5.0.3
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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

	emr-5.2.0	emr-5.1.1	emr-5.1.0	emr-5.0.3
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	0.152.3	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

	emr-5.2.0	emr-5.1.1	emr-5.1.0	emr-5.0.3
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

Notas da versão 5.2.0

As notas de lançamento a seguir incluem informações para a versão EMR 5.2.0 da Amazon. As alterações são relativas à versão EMR 5.1.0 da Amazon.

Data do release: 21 de novembro de 2016

Alterações e melhorias

- Adicionado o modo de armazenamento Amazon S3 para HBase
- Permite que você especifique uma localização do Amazon S3 para o HBase rootdir. Para obter mais informações, consulte [HBase no Amazon S3](#).

Atualizações

- Atualizado para Spark 2.0.2

Problemas conhecidos das versões anteriores que foram resolvidos

- Corrigido um problema com /mnt sendo restrito a 2 TB em tipos EBS de instância somente.
- Corrigido um problema com o controlador de instância e com os logs do logpusher serem a saída para seus arquivos .out correspondentes, em vez de para seus arquivos normais .log configurados com log4j, que mudam de hora em hora. Os arquivos .out não alternam, portanto isso acabará por encher a partição /emr. Esse problema afeta apenas os tipos de instância de máquina virtual (HVM) de hardware.

Versões de componente 5.2.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion-amzn-EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>emr-ddb</code>	4.1.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.1.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.2.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.4.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
<code>emrfs</code>	2.12.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
<code>flink-client</code>	1.1.3	Scripts do cliente da linha de comando e aplicativos do Apache Flink.

Componente	Version (Versão)	Descrição
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-0	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-0	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-0	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-0	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httpfs-server	2.7.3-amzn-0	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-0	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API

Componente	Version (Versão)	Descrição
hadoop-mapred	2.7.3-amzn-0	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-0	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-0	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-0	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.2.3	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.2.3	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.2.3	HBasecliente de linha de comando.
hbase-rest-server	1.2.3	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	1.2.3	Serviço que fornece um endpoint Thrift para. HBase

Componente	Version (Versão)	Descrição
hcatalog-client	2.1.0-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.1.0-amzn-0	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.1.0-amzn-0	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	2.1.0-amzn-0	O cliente da linha de comando do Hive.
hive-metastore-server	2.1.0-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	2.1.0-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.10.0-amzn-0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.12.2	A biblioteca de Machine Learning.
mysql-server	5.5.52	Meu servidor SQL de banco de dados.

Componente	Version (Versão)	Descrição
oozie-client	4.2.0	O cliente da linha de comando do Oozie.
oozie-server	4.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.7.0- -1,2 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.7.0- -1,2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.152.3	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.152.3	O serviço que executa partes de uma consulta.
pig-client	0.16.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	2.0.2	Os clientes da linha de comando do Spark.
spark-history-server	2.0.2	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.0.2	Mecanismo de execução na memória para YARN.

Componente	Version (Versão)	Descrição
spark-yarn-slave	2.0.2	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.23	HTTPServidor Apache.
zeppelin-server	0.6.2	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.8	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.8	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.2.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.2.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do flink-conf.yaml.
flink-log4j	Alterar as configurações de log4j.properties no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase	Configurações EMR selecionadas pela Amazon para Apache. HBase
hbase-env	Mude os valores no HBase ambiente.

Classificações	Descrição
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.

Classificações	Descrição
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.

Classificações	Descrição
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.

Classificações	Descrição
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.

Classificações	Descrição
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oracoop-site	Altere os valores no arquivo oracoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.1.1

Versões 5.1.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.1.1	emr-5.1.0	emr-5.0.3	emr-5.0.2
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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

	emr-5.1.1	emr-5.1.0	emr-5.0.3	emr-5.0.2
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	-	-	-	-
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

	emr-5.1.1	emr-5.1.0	emr-5.0.3	emr-5.0.2
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

Notas da versão 5.1.1

Esta é uma versão de patch para adicionar a autenticação AWS Signature versão 4 para solicitações ao Amazon S3. Todos os aplicativos e componentes são iguais aos da EMR versão anterior da Amazon.

Important

Nesta versão de lançamento, a Amazon EMR usa o AWS Signature versão 4 exclusivamente para autenticar solicitações para o Amazon S3. Para obter mais informações, consulte [Novidades](#).

Versões de componente 5.1.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
emr-ddb	4.1.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.1.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.2.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.4.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.11.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.1.3	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas

Componente	Version (Versão)	Descrição
		coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-0	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-0	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-0	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-0	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httpfs-server	2.7.3-amzn-0	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-0	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-0	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-0	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-0	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.

Componente	Version (Versão)	Descrição
hadoop-yarn-timeline-server	2.7.3-amzn-0	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.2.3	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.2.3	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.2.3	HBase cliente de linha de comando.
hbase-rest-server	1.2.3	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.2.3	Serviço que fornece um endpoint Thrift para HBase
hcatalog-client	2.1.0-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.1.0-amzn-0	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP endpoint fornecendo uma REST interface para HCatalog.

Componente	Version (Versão)	Descrição
hive-client	2.1.0-amzn-0	O cliente da linha de comando do Hive.
hive-metastore-server	2.1.0-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	2.1.0-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.10.0-amzn-0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.12.2	A biblioteca de Machine Learning.
mysql-server	5.5.52	Meu servidor SQL de banco de dados.
oozie-client	4.2.0	O cliente da linha de comando do Oozie.
oozie-server	4.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.7.0- -1,2 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.7.0- -1,2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API

Componente	Version (Versão)	Descrição
presto-coordinator	0.152.3	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.152.3	O serviço que executa partes de uma consulta.
pig-client	0.16.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	2.0.1	Os clientes da linha de comando do Spark.
spark-history-server	2.0.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.0.1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.0.1	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.23	HTTP Servidor Apache.

Componente	Version (Versão)	Descrição
zeppelin-server	0.6.2	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.8	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.8	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.1.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.1.1

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.

Classificações	Descrição
flink-log4j-yarn-session	Altere as configurações do Flink log4 j-yarn-session .properties.
flink-log4j-cli	Alterar as configurações de log4j-cli.properties no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.

Classificações	Descrição
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.

Classificações	Descrição
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.

Classificações	Descrição
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.

Classificações	Descrição
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.

Classificações	Descrição
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.1.0

Versões 5.1.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Flink](#), [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.1.0	emr-5.0.3	emr-5.0.2	emr-5.0.1
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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
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	-	-	-	-

	emr-5.1.0	emr-5.0.3	emr-5.0.2	emr-5.0.1
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
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

Notas da versão 5.1.0

As notas de lançamento a seguir incluem informações para a versão EMR 5.1.0 da Amazon. As alterações são relativas à versão EMR 5.0.3 da Amazon.

Data do release: 03 de novembro de 2016

Alterações e melhorias

- Adicionado o suporte para Flink 1.1.3.
- O Presto foi adicionado como uma opção na seção bloco de anotações do Hue.

Atualizações

- Atualizado para 1.2.3 HBase
- Atualizado para Zeppelin 0.6.2

Problemas conhecidos das versões anteriores que foram resolvidos

- Corrigido um problema com as consultas Tez no Amazon S3 ORC com arquivos que não funcionavam tão bem quanto nas versões anteriores do Amazon EMR 4.x.

Versões de componente 5.1.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>emr-ddb</code>	4.1.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.1.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.2.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
emr-s3-dist-cp	2.4.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.11.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
flink-client	1.1.3	Scripts do cliente da linha de comando e aplicativos do Apache Flink.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-0	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-0	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-0	HDFScliente e biblioteca de linha de comando

Componente	Version (Versão)	Descrição
hadoop-hdfs-namenode	2.7.3-amzn-0	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httfs-server	2.7.3-amzn-0	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-0	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-0	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-0	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-0	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-0	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.2.3	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.2.3	Serviço para atender a uma ou mais HBase regiões.

Componente	Version (Versão)	Descrição
hbase-client	1.2.3	HBase cliente de linha de comando.
hbase-rest-server	1.2.3	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.2.3	Serviço que fornece um endpoint Thrift para HBase
hcatalog-client	2.1.0-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.1.0-amzn-0	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.1.0-amzn-0	O cliente da linha de comando do Hive.
hive-metastore-server	2.1.0-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	2.1.0-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.

Componente	Version (Versão)	Descrição
hue-server	3.10.0-amzn-0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.12.2	A biblioteca de Machine Learning.
mysql-server	5.5.52	Meu servidor SQL de banco de dados.
oozie-client	4.2.0	O cliente da linha de comando do Oozie.
oozie-server	4.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.7.0- -1,2 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.7.0- -1,2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.152.3	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.152.3	O serviço que executa partes de uma consulta.
pig-client	0.16.0-amzn-0	O cliente da linha de comando do Pig.

Componente	Version (Versão)	Descrição
spark-client	2.0.1	Os clientes da linha de comando do Spark.
spark-history-server	2.0.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.0.1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.0.1	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.23	HTTP Servidor Apache.
zeppelin-server	0.6.2	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.8	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.8	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.1.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.1.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
flink-conf	Alterar as configurações do <code>flink-conf.yaml</code> .
flink-log4j	Alterar as configurações de <code>log4j.properties</code> no Flink.
flink-log4j-yarn-session	Altere as configurações do Flink <code>log4j-yarn-session.properties</code> .
flink-log4j-cli	Alterar as configurações de <code>log4j-cli.properties</code> no Flink.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop

Classificações	Descrição
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.

Classificações	Descrição
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.

Classificações	Descrição
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.

Classificações	Descrição
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.

Classificações	Descrição
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.0.3

Versões 5.0.3 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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

Notas da versão 5.0.3

As notas de lançamento a seguir incluem informações para a versão EMR 5.0.3 da Amazon. As alterações são relativas à versão EMR 5.0.0 da Amazon.

Data do release: 24 de outubro de 2016

Atualizações

- Atualizado para Hadoop 2.7.3
- Atualizado para Presto 0.152.3, que inclui o suporte para a interface da web do Presto. Você pode acessar a interface da web do Presto no coordenador do Presto usando a porta 8889. Para obter mais informações sobre a interface da web do Presto, consulte [Interface da web](#) na documentação do Presto.
- Atualizado para Spark 2.0.1
- Os EMR lançamentos da Amazon agora são baseados no Amazon Linux 2016.09. Para ter mais informações, consulte <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.

Versões de componente 5.0.3

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`.

O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado myapp-component com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como 2.2-amzn-2.

Componente	Version (Versão)	Descrição
emr-ddb	4.1.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.1.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.2.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.4.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.10.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas

Componente	Version (Versão)	Descrição
		coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-0	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-0	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-0	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-0	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-https-server	2.7.3-amzn-0	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-0	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-0	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-0	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-0	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.

Componente	Version (Versão)	Descrição
hadoop-yarn-timeline-server	2.7.3-amzn-0	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.2.2	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.2.2	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.2.2	HBase cliente de linha de comando.
hbase-rest-server	1.2.2	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.2.2	Serviço que fornece um endpoint Thrift para HBase.
hcatalog-client	2.1.0-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.1.0-amzn-0	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP endpoint fornecendo uma REST interface para HCatalog.

Componente	Version (Versão)	Descrição
hive-client	2.1.0-amzn-0	O cliente da linha de comando do Hive.
hive-metastore-server	2.1.0-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	2.1.0-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.10.0-amzn-0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.12.2	A biblioteca de Machine Learning.
mysql-server	5.5.52	Meu servidor SQL de banco de dados.
oozie-client	4.2.0	O cliente da linha de comando do Oozie.
oozie-server	4.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.7.0- -1,2 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.7.0- -1,2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API

Componente	Version (Versão)	Descrição
presto-coordinator	0.152.3	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.152.3	O serviço que executa partes de uma consulta.
pig-client	0.16.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	2.0.1	Os clientes da linha de comando do Spark.
spark-history-server	2.0.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.0.1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.0.1	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.23	HTTP Servidor Apache.

Componente	Version (Versão)	Descrição
zeppelin-server	0.6.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.8	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.8	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.0.3

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.0.3

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Hadoop.

Classificações	Descrição
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.

Classificações	Descrição
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.

Classificações	Descrição
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.

Classificações	Descrição
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.

Classificações	Descrição
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.

Classificações	Descrição
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.0.2

Versões 5.0.2 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
Scala	2.11.8	2.11.8	2.11.8	2.11.8
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
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
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

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
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

Notas da versão 5.0.2

Versões de componente 5.0.2

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
emr-ddb	4.0.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.1.0	Bibliotecas convenientes para o ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
emr-kinesis	3.2.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.4.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.9.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.2-amzn-3	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.2-amzn-3	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.2-amzn-3	HDFScliente e biblioteca de linha de comando

Componente	Version (Versão)	Descrição
hadoop-hdfs-namenode	2.7.2-amzn-3	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httfs-server	2.7.2-amzn-3	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.2-amzn-3	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.2-amzn-3	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.2-amzn-3	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.2-amzn-3	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.2-amzn-3	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.2.2	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.2.2	Serviço para atender a uma ou mais HBase regiões.

Componente	Version (Versão)	Descrição
hbase-client	1.2.2	HBase cliente de linha de comando.
hbase-rest-server	1.2.2	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.2.2	Serviço que fornece um endpoint Thrift para HBase.
hcatalog-client	2.1.0-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.1.0-amzn-0	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.1.0-amzn-0	O cliente da linha de comando do Hive.
hive-metastore-server	2.1.0-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	2.1.0-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.

Componente	Version (Versão)	Descrição
hue-server	3.10.0-amzn-0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.12.2	A biblioteca de Machine Learning.
mysql-server	5.5.46	Meu servidor SQL de banco de dados.
oozie-client	4.2.0	O cliente da linha de comando do Oozie.
oozie-server	4.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.7.0- -1,2 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.7.0- -1,2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.150	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.150	O serviço que executa partes de uma consulta.
pig-client	0.16.0-amzn-0	O cliente da linha de comando do Pig.

Componente	Version (Versão)	Descrição
spark-client	2.0.0	Os clientes da linha de comando do Spark.
spark-history-server	2.0.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.0.0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.0.0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.23	HTTP Servidor Apache.
zeppelin-server	0.6.1- SNAPSHOT	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.8	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.8	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.0.2

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.0.2

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase <code>hbase-log4j.properties</code> .
hbase-metrics	Altere os valores no arquivo HBase <code>hadoop-metrics2-hbase.properties</code> do.
hbase-policy	Altere os valores no HBase arquivo <code>hbase-policy.xml</code> do.

Classificações	Descrição
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.

Classificações	Descrição
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.

Classificações	Descrição
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.

Classificações	Descrição
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.

Classificações	Descrição
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.0.1

Versões 5.0.1 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)

- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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	-	-	-	-

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
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
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

Notas da versão 5.0.1

Versões de componente 5.0.1

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são

a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion-amzn-EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado myapp-component com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como 2.2-amzn-2.

Componente	Version (Versão)	Descrição
emr-ddb	4.0.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.1.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.2.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.4.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.9.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.

Componente	Version (Versão)	Descrição
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.2-amzn-3	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.2-amzn-3	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.2-amzn-3	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.2-amzn-3	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-https-server	2.7.2-amzn-3	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.2-amzn-3	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.2-amzn-3	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.2-amzn-3	YARNserviço para gerenciar contêineres em um nó individual.

Componente	Version (Versão)	Descrição
hadoop-yarn-resourcemanager	2.7.2-amzn-3	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.2-amzn-3	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.2.2	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.2.2	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.2.2	HBasecliente de linha de comando.
hbase-rest-server	1.2.2	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	1.2.2	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	2.1.0-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.1.0-amzn-0	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.

Componente	Version (Versão)	Descrição
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.1.0-amzn-0	O cliente da linha de comando do Hive.
hive-metastore-server	2.1.0-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	2.1.0-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.10.0-amzn-0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.12.2	A biblioteca de Machine Learning.
mysql-server	5.5.46	Meu servidor SQL de banco de dados.
oozie-client	4.2.0	O cliente da linha de comando do Oozie.
oozie-server	4.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.7.0- -1,2 HBase	As bibliotecas do phoenix para servidor e cliente

Componente	Version (Versão)	Descrição
phoenix-query-server	4.7.0- -1,2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.150	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.150	O serviço que executa partes de uma consulta.
pig-client	0.16.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	2.0.0	Os clientes da linha de comando do Spark.
spark-history-server	2.0.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.0.0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.0.0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.

Componente	Version (Versão)	Descrição
webserver	2.4.23	HTTPServidor Apache.
zeppelin-server	0.6.1- SNAPSHOT	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.8	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.8	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.0.1

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.0.1

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.

Classificações	Descrição
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.

Classificações	Descrição
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.

Classificações	Descrição
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.

Classificações	Descrição
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.

Classificações	Descrição
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.

Classificações	Descrição
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 5.0.0

Versões 5.0.0 da aplicação

As seguintes aplicações são compatíveis com esta versão: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie](#), [Phoenix](#), [Pig](#), [Presto](#), [Spark](#), [Sqoop](#), [Tez](#), [Zeppelin](#) e [ZooKeeper](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
Scala	2.11.8	2.11.8	2.11.8	2.11.8

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
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
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

	emr-5.0.3	emr-5.0.2	emr-5.0.1	emr-5.0.0
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

Notas da versão 5.0.0

Data do release: 27 de julho de 2016

Atualizações

- Atualizado para Hive 2.1
- Atualizado para Presto 0.150
- Atualizado para Spark 2.0
- Atualizado para Hue 3.10.0
- Atualizado para Pig 0.16.0
- Atualizado para Tez 0.8.4
- Atualizado para Zeppelin 0.6.1

Alterações e melhorias

- A Amazon EMR oferece suporte às versões mais recentes de código aberto do Hive (versão 2.1) e do Pig (versão 0.16.0). Se você já usou o Hive ou o Pig EMR na Amazon no passado, isso pode afetar alguns casos de uso. Para obter mais informações, consulte [Hive](#) e [Pig](#).

- O mecanismo de execução padrão para o Hive e o Pig agora é o Tez. Para alterar isso, você deve editar os valores apropriados nas classificações de configuração `hive-site` e `pig-properties`, respectivamente.
- Um recurso de etapa aprimorada de depuração foi adicionado, o que permite que você veja a causa raiz de falhas de etapa se o serviço puder determinar a causa. Para obter mais informações, consulte [Enhanced Step Debugging](#) no Amazon EMR Management Guide.
- Os aplicativos que, anteriormente, terminavam com "-Sandbox" não têm mais esse sufixo. Isso pode inutilizar sua automação, por exemplo, se você estiver usando scripts para iniciar clusters com esses aplicativos. A tabela a seguir mostra os nomes dos aplicativos no Amazon EMR 4.7.2 versus no Amazon EMR 5.0.0.

Alterações dos nomes de aplicativos

Amazon EMR 4.7.2	Amazon EMR 5.0.0
Oozie-Sandbox	Oozie
Presto-Sandbox	Presto
Sqoop-Sandbox	Sqoop
Zeppelin-Sandbox	Zeppelin
ZooKeeper-Caixa de areia	ZooKeeper

- O Spark agora está compilado para Scala 2.11.
- O Java 8 agora é o padrão JVM. Todas as aplicações são executadas usando o runtime do Java 8. Não há alterações em qualquer destino de código de bytes da aplicação. A maioria dos aplicativos continuam a usar o Java 7 como destino.
- O Zeppelin agora inclui recursos de autenticação. Para obter mais informações, consulte [Zeppelin](#).
- Adicionado o suporte para configurações de segurança, que permitem criar e aplicar opções de criptografia com mais facilidade. Para obter mais informações, consulte [Criptografia de dados](#).

Versões de componente 5.0.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou

aws. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado myapp-component com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como 2.2-amzn-2.

Componente	Version (Versão)	Descrição
emr-ddb	4.0.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.1.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.2.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.4.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.9.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas

Componente	Version (Versão)	Descrição
		de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.2-amzn-3	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.2-amzn-3	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.2-amzn-3	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.2-amzn-3	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httpfs-server	2.7.2-amzn-3	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.2-amzn-3	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.2-amzn-3	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.2-amzn-3	YARNserviço para gerenciar contêineres em um nó individual.

Componente	Version (Versão)	Descrição
hadoop-yarn-resourcemanager	2.7.2-amzn-3	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.2-amzn-3	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.2.2	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.2.2	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.2.2	HBasecliente de linha de comando.
hbase-rest-server	1.2.2	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	1.2.2	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	2.1.0-amzn-0	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	2.1.0-amzn-0	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.

Componente	Version (Versão)	Descrição
hcatalog-webhcat-server	2.1.0-amzn-0	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	2.1.0-amzn-0	O cliente da linha de comando do Hive.
hive-metastore-server	2.1.0-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	2.1.0-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.10.0-amzn-0	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.12.2	A biblioteca de Machine Learning.
mysql-server	5.5.46	Meu servidor SQL de banco de dados.
oozie-client	4.2.0	O cliente da linha de comando do Oozie.
oozie-server	4.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.7.0- -1,2 HBase	As bibliotecas do phoenix para servidor e cliente

Componente	Version (Versão)	Descrição
phoenix-query-server	4.7.0- -1,2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.150	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.150	O serviço que executa partes de uma consulta.
pig-client	0.16.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	2.0.0	Os clientes da linha de comando do Spark.
spark-history-server	2.0.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	2.0.0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	2.0.0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.

Componente	Version (Versão)	Descrição
webserver	2.4.23	HTTPServidor Apache.
zeppelin-server	0.6.1- SNAPSHOT	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.8	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.8	ZooKeeper cliente de linha de comando.

Classificações de configuração 5.0.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-5.0.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.

Classificações	Descrição
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.

Classificações	Descrição
hcatalog-webhcat-log4j2	Altere os valores nas propriedades ebHCat log4j2.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-beeline-log4j2	Alterar os valores no arquivo beeline-log4j2.properties do Hive.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j2	Altere os valores no arquivo hive-exec-log4j2.properties do Hive.
hive-llap-daemon-log4j2	Altere os valores no arquivo llap-daemon-log4j2.properties do Hive.
hive-log4j2	Alterar os valores no arquivo hive-log4j2.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.

Classificações	Descrição
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.

Classificações	Descrição
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.

Classificações	Descrição
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-hive-site	Alterar os valores no arquivo hive-site.xml do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.

Classificações	Descrição
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Versões de lançamento do Amazon EMR 4.x

Esta seção contém versões de aplicativos, notas de lançamento, versões de componentes e classificações de configuração disponíveis em cada versão de lançamento do Amazon EMR 4.x.

Ao iniciar um cluster, você pode escolher entre várias versões da AmazonEMR. Isso permite que você teste e use versões de aplicações que atendem aos seus requisitos de compatibilidade. Você especifica o número da versão com o rótulo da versão. Os rótulos de versão estão no formato `emr-x.x.x`. Por exemplo, `emr-7.2.0`.

Novos EMR lançamentos da Amazon são disponibilizados em diferentes regiões por um período de vários dias, começando com a primeira região na data de lançamento inicial. A versão mais recente pode não estar disponível em sua região durante esse período.

Para obter uma tabela abrangente das versões do aplicativo em cada versão EMR 4.x da Amazon, consulte [Versões de aplicações nas versões 4.x do Amazon EMR](#).

Tópicos

- [Versões de aplicações nas versões 4.x do Amazon EMR](#)
- [Diferenças nas versões 4.x do Amazon EMR](#)
- [Amazon EMR versão 4.9.6](#)
- [Amazon EMR versão 4.9.5](#)
- [Amazon EMR versão 4.9.4](#)
- [Amazon EMR versão 4.9.3](#)
- [Amazon EMR versão 4.9.2](#)
- [Amazon EMR versão 4.9.1](#)
- [Amazon EMR versão 4.8.5](#)
- [Amazon EMR versão 4.8.4](#)

- [Amazon EMR versão 4.8.3](#)
- [Amazon EMR versão 4.8.2](#)
- [Amazon EMR versão 4.8.1](#)
- [Amazon EMR versão 4.8.0](#)
- [Amazon EMR versão 4.7.4](#)
- [Amazon EMR versão 4.7.3](#)
- [Amazon EMR versão 4.7.2](#)
- [Amazon EMR versão 4.7.1](#)
- [Amazon EMR versão 4.7.0](#)
- [Amazon EMR versão 4.6.1](#)
- [Amazon EMR versão 4.6.0](#)
- [Amazon EMR versão 4.5.0](#)
- [Amazon EMR versão 4.4.0](#)
- [Amazon EMR versão 4.3.0](#)
- [Amazon EMR versão 4.2.0](#)
- [Amazon EMR versão 4.1.0](#)
- [Amazon EMR versão 4.0.0](#)

Versões de aplicações nas versões 4.x do Amazon EMR

Para obter uma tabela abrangente que liste as versões das aplicações disponíveis em cada versão 4.x do Amazon EMR, abra as [versões de aplicações nas versões 4.x do Amazon EMR](#) no seu navegador.

Diferenças nas versões 4.x do Amazon EMR

A documentação para atributos do Amazon EMR no Guia de gerenciamento do Amazon EMR especifica a versão do Amazon EMR na qual um atributo foi disponibilizado, além das diferenças aplicáveis entre atributos do Amazon EMR desde a versão 4.0.0.

Desde a versão 5.0.0 do Amazon EMR, algumas aplicações receberam uma atualização de versão significativa, que alterou detalhes de instalação ou operacionais e outras foram promovidas de aplicações em sandbox a aplicações nativas. Cada tópico desta seção apresenta diferenças significativas e específicas das aplicações quando são usadas as versões 4.x do Amazon EMR.

Tópicos

- [Aplicações em sandbox](#)
- [Considerações sobre o uso do Hive no Amazon EMR 4.x](#)
- [Considerações sobre o uso do Pig no Amazon EMR 4.x](#)

Aplicações em sandbox

Ao serem usadas as versões 4.x do Amazon EMR, algumas aplicações são consideradas aplicações em sandbox. As aplicações em sandbox são versões anteriores da aplicação que disponibilizamos no momento da versão inicial do Amazon EMR devido à demanda. Você pode usar o console, a AWS CLI ou a API para que o Amazon EMR instale aplicações em sandbox da mesma maneira que aplicações nativas, mas as aplicações em sandbox têm suporte e documentação limitados. As aplicações em Sandbox se tornaram aplicações nativas, totalmente compatíveis com as versões 5.0.0 e posteriores do Amazon EMR. Veja a seguir as aplicações em sandbox nas versões 4.x do Amazon EMR:

- Oozie
- Presto
- Sqoop
- Zeppelin
- ZooKeeper

Quando você instala aplicativos em sandbox, os nomes dos aplicativos são representados pelo sufixo `-sandbox`. Por exemplo, para instalar a versão em sandbox do *Presto*, use `Presto-sandbox`. A instalação pode demorar mais do que para um aplicativo totalmente compatível. Os números de versão listados para cada aplicativo nesta seção correspondem à versão da comunidade do aplicativo.

Oozie (versões em sandbox)

O Oozie está disponível como uma aplicação em sandbox desde a versão 4.1.0 do Amazon EMR.

Os exemplos do Oozie não são instalados por padrão usando as versões em sandbox. Para instalar os exemplos, execute o SSH no nó principal do cluster e `install-oozie-examples`.

Informações sobre a versão em sandbox do Oozie

Rótulo de versão do Amazon EMR	Oozie (versão em sandbox)	Componentes instalados com o 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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, 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-resourcemanager, 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-resour

Rótulo de versão do Amazon EMR	Oozie (versão em sandbox)	Componentes instalados com o Oozie-Sandbox
		cemanager, oozie-client, oozie-server
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-resourcecemanager, 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-resourcecemanager, oozie-client, oozie-server

Rótulo de versão do Amazon EMR	Oozie (versão em sandbox)	Componentes instalados com o 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-resource-manager, 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-resource-manager, oozie-client, oozie-server

Rótulo de versão do Amazon EMR	Oozie (versão em sandbox)	Componentes instalados com o 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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, 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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server

Rótulo de versão do Amazon EMR	Oozie (versão em sandbox)	Componentes instalados com o 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-resource-manager, 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-resource-manager, oozie-client, oozie-server

Rótulo de versão do Amazon EMR	Oozie (versão em sandbox)	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Oozie (versão em sandbox)	Componentes instalados com o 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-https-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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, oozie-client, oozie-server

Rótulo de versão do Amazon EMR	Oozie (versão em sandbox)	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Oozie (versão em sandbox)	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Oozie (versão em sandbox)	Componentes instalados com o 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 (versões em sandbox)

O Presto está disponível como aplicação em sandbox desde a versão 4.1.0 do Amazon EMR.

Informações sobre a versão do Presto-Sandbox

Rótulo de versão do Amazon EMR	Presto (versão em sandbox)	Componentes instalados com o 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, hcatalog-server, mysql-server,

Rótulo de versão do Amazon EMR	Presto (versão em sandbox)	Componentes instalados com o Presto-Sandbox
		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

Rótulo de versão do Amazon EMR	Presto (versão em sandbox)	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Presto (versão em sandbox)	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Presto (versão em sandbox)	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Presto (versão em sandbox)	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Presto (versão em sandbox)	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Presto (versão em sandbox)	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Presto (versão em sandbox)	Componentes instalados com o 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 (versões em sandbox)

O Sqoop está disponível como uma aplicação em sandbox desde a versão 4.4.0 do Amazon EMR.

Informações sobre a versão do Sqoop-Sandbox

Rótulo de versão do Amazon EMR	Sqoop (versão em sandbox)	Componentes instalados com o 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-https-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-hdfs-namenode, hadoop-

Rótulo de versão do Amazon EMR	Sqoop (versão em sandbox)	Componentes instalados com o Sqoop-Sandbox
		httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client
emr-4.9.4	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
emr-4.9.3	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

Rótulo de versão do Amazon EMR	Sqoop (versão em sandbox)	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Sqoop (versão em sandbox)	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Sqoop (versão em sandbox)	Componentes instalados com o 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-https-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-https-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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Rótulo de versão do Amazon EMR	Sqoop (versão em sandbox)	Componentes instalados com o 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-httpfs-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-httpfs-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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Rótulo de versão do Amazon EMR	Sqoop (versão em sandbox)	Componentes instalados com o 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-httpfs-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-httpfs-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-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client

Rótulo de versão do Amazon EMR	Sqoop (versão em sandbox)	Componentes instalados com o 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 (versões em sandbox)

O Zeppelin está disponível como uma aplicação em sandbox desde a versão 4.1.0 do Amazon EMR.

Informações sobre a versão do Zeppelin-Sandbox

Rótulo de versão do Amazon EMR	Zeppelin (versão em sandbox)	Componentes instalados com o 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-server, hadoop-yarn-nodema

Rótulo de versão do Amazon EMR	Zeppelin (versão em sandbox)	Componentes instalados com o Zeppelin-Sandbox
		nager, hadoop-yarn-resourcemanager, 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-resourcemanager, 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-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Rótulo de versão do Amazon EMR	Zeppelin (versão em sandbox)	Componentes instalados com o Zeppelin-Sandbox
emr-4.9.3	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-4.9.2	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

Rótulo de versão do Amazon EMR	Zeppelin (versão em sandbox)	Componentes instalados com o Zeppelin-Sandbox
emr-4.9.1	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-4.8.5	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

Rótulo de versão do Amazon EMR	Zeppelin (versão em sandbox)	Componentes instalados com o Zeppelin-Sandbox
emr-4.8.4	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-resourcemanager, 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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Rótulo de versão do Amazon EMR	Zeppelin (versão em sandbox)	Componentes instalados com o Zeppelin-Sandbox
emr-4.8.2	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-resourcemanager, 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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Rótulo de versão do Amazon EMR	Zeppelin (versão em sandbox)	Componentes instalados com o Zeppelin-Sandbox
emr-4.8.0	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-resourcemanager, 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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Rótulo de versão do Amazon EMR	Zeppelin (versão em sandbox)	Componentes instalados com o Zeppelin-Sandbox
emr-4.7.3	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-4.7.2	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

Rótulo de versão do Amazon EMR	Zeppelin (versão em sandbox)	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Zeppelin (versão em sandbox)	Componentes instalados com o 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-ya rn-resourcemanager, 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-ya rn-resourcemanager, 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-ya rn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server

Rótulo de versão do Amazon EMR	Zeppelin (versão em sandbox)	Componentes instalados com o 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 (versões em sandbox)

O Zookeeper está disponível como uma aplicação em sandbox desde a versão 4.6.0 do Amazon EMR.

Informações sobre a versão do ZooKeeper-Sandbox

Rótulo de versão do Amazon EMR	Zookeeper (versão em sandbox)	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Zookeeper (versão em sandbox)	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, zookeeper-client, zookeeper-server

Rótulo de versão do Amazon EMR	Zookeeper (versão em sandbox)	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Zookeeper (versão em sandbox)	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Zookeeper (versão em sandbox)	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Zookeeper (versão em sandbox)	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Zookeeper (versão em sandbox)	Componentes instalados com o 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

Considerações sobre o uso do Hive no Amazon EMR 4.x

Esta seção discute as diferenças que devem ser consideradas ao ser usada a versão 1.0.0 do Hive nas versões 4.x do Amazon EMR em comparação com o Hive 2.x nas versões 5.x do Amazon EMR.

Transações ACID não são compatíveis

O Hive nas versões 4.x do Amazon EMR não é compatível com transações ACID com dados do Hive armazenados no Amazon S3 ao serem usadas versões 4.x. Se você tentar criar uma tabela transacional no Amazon S3, ocorrerá uma exceção.

Leitura e gravação em tabelas no Amazon S3

O Hive nas versões 4.x do Amazon EMR pode gravar diretamente no Amazon S3 sem o uso de arquivos temporários. Isso melhora a performance, mas uma consequência é que você não pode ler e gravar na mesma tabela no Amazon S3 dentro da mesma instrução do Hive. Uma solução é criar e usar uma tabela temporária no HDFS.

O exemplo a seguir mostra como usar várias instruções do Hive para atualizar uma tabela no Amazon S3. As instruções criam uma tabela temporária no HDFS denominada tmp com base em uma tabela no Amazon S3 denominada my_s3_table. A tabela no Amazon S3 é atualizada com o conteúdo da tabela temporária.

```
CREATE TEMPORARY TABLE tmp LIKE my_s3_table;
INSERT OVERWRITE TABLE tmp SELECT ....;
INSERT OVERWRITE TABLE my_s3_table SELECT * FROM tmp;
```

Log4j vs. Log4j 2

O Hive nas versões 4.x do Amazon EMR usa o Log4j. O padrão é começar com o Log4j 2 versão 5.0.0. Essas versões podem exigir diferentes configurações de registro. Para obter detalhes, consulte [Apache Log4j 2](#).

O MapReduce é o mecanismo de execução padrão

O Hive nas versões 4.x do Amazon EMR usa o MapReduce como mecanismo de execução padrão. Desde a versão 5.0.0 do Amazon EMR, o Tez é o padrão, fornecendo melhor performance para a maioria dos fluxos de trabalho.

Autorização do Hive

O Hive nas versões 4.x do Amazon EMR é compatível com a [autorização do Hive](#) para o HDFS, mas não para o EMRFS e o Amazon S3. Os clusters do Amazon EMR são executados com a autorização desabilitada por padrão.

Comportamento de mesclagem de arquivos do Hive com o Amazon S3

O Hive nas versões 4.x do Amazon EMR mescla arquivos pequenos no final de um trabalho somente mapa se `hive.merge.mapfiles` é `true`. A mesclagem apenas será acionada se o tamanho médio de saída do trabalho for menor do que a configuração de `hive.merge.smallfiles.avgsize`. O Hive no Amazon EMR tem exatamente o mesmo comportamento se o caminho de saída final está no HDFS. Porém, se o caminho de saída estiver no Amazon S3, o parâmetro `hive.merge.smallfiles.avgsize` será ignorado. Nessa situação, a tarefa de mesclagem sempre será acionada se `hive.merge.mapfiles` estiver definido como `true`.

Considerações sobre o uso do Pig no Amazon EMR 4.x

O Pig versão 0.14.0 é instalado em clusters criados usando versões 4.x do Amazon EMR. O Pig foi atualizado para a versão 0.16.0 no Amazon EMR 5.0.0. As diferenças significativas são abordadas abaixo.

Mecanismo de execução padrão diferente

O Pig versão 0.14.0 nas versões 4.x do Amazon EMR usa o MapReduce como mecanismo de execução padrão. O Pig 0.16.0 e versões posteriores usam o Apache Tez. Você pode definir

explicitamente `execType=mapreduce` na classificação de configuração `pig-properties` para usar o MapReduce.

Funções definidas pelo usuário (UDFs) do Pig descartadas

UDFs personalizadas que estavam disponíveis nas versões 4.x do Amazon EMR foram descartadas desde o Pig 0.16.0. A maioria das UDFs tem funções equivalentes que você pode usar em vez disso. A tabela a seguir lista UDFs descartadas e funções equivalentes. Para obter mais informações, consulte [Funções integradas](#) no site do Apache Pig.

UDF descartada	Função equivalente
FORMAT_DT(dtformat, date)	GetHour(date), GetMinute(date), GetMonth(date), GetSecond(date), GetWeek(date), GetYear(date), GetDay(date)
EXTRACT(string, padrão)	REGEX_EXTRACT_ALL(string, pattern)
REPLACE(string, pattern, replacement)	REPLACE(string, pattern, replacement)
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, duration)	AddDuration(data, duração), SubtractDuration(data, duração)
PERIOD(dt, dt2)	WeeksBetween(dt, dt2), YearsBetween(dt, dt2), SecondsBetween(dt, dt2), MonthsBetween(dt, dt2), MinutesBetween(dt, dt2), HoursBetween(dt, dt2)
CAPITALIZE(string)	UCFIRST(string)

UDF descartada	Função equivalente
CONCAT_WITH()	CONCAT()
INDEX_OF()	INDEXOF()
LAST_INDEX_OF()	LAST_INDEXOF()
SPLIT_ON_REGEX()	STRSPLT()
UNCAPITALIZE()	LCFIRST()

As seguintes UDFs foram descartadas sem equivalente: FORMAT(), LOCAL_DATE(), LOCAL_TIME(), CENTER(), LEFT_PAD(), REPEAT(), REPLACE_ONCE(), RIGHT_PAD(), STRIP(), STRIP_END(), STRIP_START(), SWAP_CASE().

Comandos Grunt descontinuados

Alguns comandos Grunt foram descontinuados a partir do Pig 0.16.0. A tabela a seguir lista os comandos Grunt no Pig 0.14.0 e os comandos equivalentes na versão atual, quando aplicável.

Comandos Grunt do Pig 0.14.0 e seus equivalentes atuais

Comando Grunt do Pig 0.14.0	Comando Grunt do Pig 0.16.0 e versões posteriores
cat <non-hdfs-path>)	fs -cat <non-hdfs-path>;
cd <non-hdfs-path>;	Não há equivalente
ls <non-hdfs-path>;	fs -ls <non-hdfs-path>;
move <non-hdfs-path> <non-hdfs-path>;	fs -mv <non-hdfs-path> <non-hdfs-path>;
copy <non-hdfs-path> <non-hdfs-path>;	fs -cp <non-hdfs-path> <non-hdfs-path>;
copyToLocal <non-hdfs-path> <local-path>;	fs -copyToLocal <non-hdfs-path> <local-path>;
copyFromLocal <local-path> <non-hdfs-path>;	fs -copyFromLocal <local-path> <non-hdfs-path>;

Comando Grunt do Pig 0.14.0	Comando Grunt do Pig 0.16.0 e versões posteriores
<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>

Capacidade removida para diretórios iniciais não HDFS

O Pig 0.14.0 nas versões 4.x do Amazon EMR tem dois mecanismos para permitir a execução de scripts do Pig por usuários que não sejam o usuário do hadoop e não tenham diretórios iniciais. O primeiro mecanismo é um fallback automático que define o diretório de trabalho inicial como o diretório raiz caso o diretório inicial não exista. O segunda é uma propriedade `pig.initial.fs.name` que permite alterar o diretório de trabalho inicial.

Esses mecanismos não estão disponíveis desde a versão 5.0.0 do Amazon EMR e os usuários devem ter um diretório inicial no HDFS. Isso não se aplica ao usuário hadoop, pois um diretório inicial é provisionado na execução. Os scripts executados usando etapas JAR do Hadoop assumem como padrão o usuário do Hadoop, a menos que outro usuário seja especificado explicitamente com o uso de `command-runner.jar`.

Amazon EMR versão 4.9.6

4.9.6 versões do aplicativo

As seguintes aplicações são compatíveis com esta versão: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) e [ZooKeeper-Sandbox](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)

- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-4.9.6	emr-4.9.5	emr-4.9.4	emr-4.9.3
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
Scala	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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	-	-	-	-

	emr-4.9.6	emr-4.9.5	emr-4.9.4	emr-4.9.3
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	-	-	-	-
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

Notas de lançamento do 4.9.6

Esta é uma versão de patch para adicionar a autenticação AWS Signature versão 4 para solicitações ao Amazon S3. Todos os aplicativos e componentes são iguais aos da EMR versão anterior da Amazon.

Important

Nesta versão de lançamento, a Amazon EMR usa o AWS Signature versão 4 exclusivamente para autenticar solicitações para o Amazon S3. Para obter mais informações, consulte [Novidades](#).

4.9.6 versões de componentes

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>emr-ddb</code>	4.3.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.2.0	Bibliotecas convenientes para o ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
emr-kinesis	3.3.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.4.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.17.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-2	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-2	HDFScliente e biblioteca de linha de comando

Componente	Version (Versão)	Descrição
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httfs-server	2.7.3-amzn-2	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-2	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.2.2	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.2.2	Serviço para atender a uma ou mais HBase regiões.

Componente	Version (Versão)	Descrição
hbase-client	1.2.2	HBase cliente de linha de comando.
hbase-rest-server	1.2.2	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.2.2	Serviço que fornece um endpoint Thrift para HBase.
hcatalog-client	1.0.0-amzn-9	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	1.0.0-amzn-9	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	1.0.0-amzn-9	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	1.0.0-amzn-9	O cliente da linha de comando do Hive.
hive-metastore-server	1.0.0-amzn-9	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	1.0.0-amzn-9	O serviço que aceita as consultas do Hive como solicitações da web.

Componente	Version (Versão)	Descrição
hue-server	3.7.1-amzn-7	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.12.2	A biblioteca de Machine Learning.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
oozie-client	4.2.0	O cliente da linha de comando do Oozie.
oozie-server	4.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.7.0- -1,2 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.7.0- -1,2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.157.1	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.157.1	O serviço que executa partes de uma consulta.
pig-client	0.14.0-amzn-0	O cliente da linha de comando do Pig.

Componente	Version (Versão)	Descrição
spark-client	1.6.3	Os clientes da linha de comando do Spark.
spark-history-server	1.6.3	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	1.6.3	Mecanismo de execução na memória para YARN.
spark-yarn-slave	1.6.3	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.
zeppelin-server	0.6.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.9	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.9	ZooKeeper cliente de linha de comando.

4.9.6 classificações de configuração

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-4.9.6

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase <code>hbase-log4j.properties</code> .
hbase-metrics	Altere os valores no arquivo HBase <code>hadoop-metrics2-hbase.properties</code> do.
hbase-policy	Altere os valores no HBase arquivo <code>hbase-policy.xml</code> do.

Classificações	Descrição
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j	Altere os valores nas propriedades ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j	Altere os valores no arquivo hive-exec-log4j.properties do Hive.
hive-log4j	Alterar os valores no arquivo hive-log4j.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue

Classificações	Descrição
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.

Classificações	Descrição
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.

Classificações	Descrição
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.

Classificações	Descrição
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 4.9.5

4.9.5 versões do aplicativo

As seguintes aplicações são compatíveis com esta versão: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) e [ZooKeeper-Sandbox](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-4.9.5	emr-4.9.4	emr-4.9.3	emr-4.9.2
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75

	emr-4.9.5	emr-4.9.4	emr-4.9.3	emr-4.9.2
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
Scala	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-

	emr-4.9.5	emr-4.9.4	emr-4.9.3	emr-4.9.2
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

Notas de lançamento do 4.9.5

As notas de lançamento a seguir incluem informações sobre a EMR versão 4.9.5 da Amazon. As alterações são referentes à versão 4.9.4.

Data da versão inicial: 29 de agosto de 2018

Alterações, melhorias e problemas resolvidos

- HBase
 - Esta versão aborda uma possível vulnerabilidade de segurança.

4.9.5 versões de componentes

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-`amzn-EmrVersion`. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>emr-ddb</code>	4.3.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.2.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.3.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.4.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.

Componente	Version (Versão)	Descrição
emrfs	2.17.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-2	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-2	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httpfs-server	2.7.3-amzn-2	HTTPendpoint para HDFS operações.

Componente	Version (Versão)	Descrição
hadoop-kms-server	2.7.3-amzn-2	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.2.2	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.2.2	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.2.2	HBasecliente de linha de comando.
hbase-rest-server	1.2.2	Serviço que fornece um RESTful HTTP endpoint paraHBase.

Componente	Version (Versão)	Descrição
hbase-thrift-server	1.2.2	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	1.0.0-amzn-9	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	1.0.0-amzn-9	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	1.0.0-amzn-9	HTTPendpoint fornecendo o uma REST interface paraHCatalog.
hive-client	1.0.0-amzn-9	O cliente da linha de comando do Hive.
hive-metastore-server	1.0.0-amzn-9	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	1.0.0-amzn-9	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.7.1-amzn-7	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.12.2	A biblioteca de Machine Learning.

Componente	Version (Versão)	Descrição
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
oozie-client	4.2.0	O cliente da linha de comando do Oozie.
oozie-server	4.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.7.0- -1,2 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.7.0- -1,2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.157.1	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.157.1	O serviço que executa partes de uma consulta.
pig-client	0.14.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	1.6.3	Os clientes da linha de comando do Spark.
spark-history-server	1.6.3	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.

Componente	Version (Versão)	Descrição
spark-on-yarn	1.6.3	Mecanismo de execução na memória para YARN.
spark-yarn-slave	1.6.3	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.
zeppelin-server	0.6.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.9	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.9	ZooKeeper cliente de linha de comando.

4.9.5 classificações de configuração

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-4.9.5

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.

Classificações	Descrição
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j	Altere os valores nas propriedades ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j	Altere os valores no arquivo hive-exec-log4j.properties do Hive.
hive-log4j	Alterar os valores no arquivo hive-log4j.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.

Classificações	Descrição
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.

Classificações	Descrição
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.

Classificações	Descrição
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.

Classificações	Descrição
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 4.9.4

4.9.4 versões do aplicativo

As seguintes aplicações são compatíveis com esta versão: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) e [ZooKeeper-Sandbox](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-4.9.4	emr-4.9.3	emr-4.9.2	emr-4.9.1
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas

	emr-4.9.4	emr-4.9.3	emr-4.9.2	emr-4.9.1
Scala	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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	-	-	-	-
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

	emr-4.9.4	emr-4.9.3	emr-4.9.2	emr-4.9.1
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

Notas de lançamento do 4.9.4

As notas de lançamento a seguir incluem informações sobre a EMR versão 4.9.4 da Amazon. As alterações são referentes à versão 4.9.3.

Data da versão inicial: 29 de março de 2018

Alterações, melhorias e problemas resolvidos

- Atualizou o kernel Amazon Linux do default Amazon Linux AMI para Amazon EMR para resolver possíveis vulnerabilidades.

4.9.4 versões de componentes

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>emr-ddb</code>	4.3.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.2.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.3.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.4.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.

Componente	Version (Versão)	Descrição
emrfs	2.17.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-2	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-2	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httpfs-server	2.7.3-amzn-2	HTTPendpoint para HDFS operações.

Componente	Version (Versão)	Descrição
hadoop-kms-server	2.7.3-amzn-2	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.2.2	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.2.2	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.2.2	HBasecliente de linha de comando.
hbase-rest-server	1.2.2	Serviço que fornece um RESTful HTTP endpoint paraHBase.

Componente	Version (Versão)	Descrição
hbase-thrift-server	1.2.2	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	1.0.0-amzn-9	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	1.0.0-amzn-9	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	1.0.0-amzn-9	HTTPendpoint fornecendo o uma REST interface paraHCatalog.
hive-client	1.0.0-amzn-9	O cliente da linha de comando do Hive.
hive-metastore-server	1.0.0-amzn-9	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	1.0.0-amzn-9	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.7.1-amzn-7	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.12.2	A biblioteca de Machine Learning.

Componente	Version (Versão)	Descrição
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
oozie-client	4.2.0	O cliente da linha de comando do Oozie.
oozie-server	4.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.7.0- -1,2 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.7.0- -1,2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.157.1	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.157.1	O serviço que executa partes de uma consulta.
pig-client	0.14.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	1.6.3	Os clientes da linha de comando do Spark.
spark-history-server	1.6.3	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.

Componente	Version (Versão)	Descrição
spark-on-yarn	1.6.3	Mecanismo de execução na memória para YARN.
spark-yarn-slave	1.6.3	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.
zeppelin-server	0.6.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.9	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.9	ZooKeeper cliente de linha de comando.

4.9.4 classificações de configuração

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-4.9.4

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.

Classificações	Descrição
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j	Altere os valores nas propriedades ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j	Altere os valores no arquivo hive-exec-log4j.properties do Hive.
hive-log4j	Alterar os valores no arquivo hive-log4j.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.

Classificações	Descrição
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.

Classificações	Descrição
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.

Classificações	Descrição
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.

Classificações	Descrição
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 4.9.3

4.9.3 versões do aplicativo

As seguintes aplicações são compatíveis com esta versão: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) e [ZooKeeper-Sandbox](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-4.9.3	emr-4.9.2	emr-4.9.1	emr-4.8.5
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas

	emr-4.9.3	emr-4.9.2	emr-4.9.1	emr-4.8.5
Scala	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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	-	-	-	-
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

	emr-4.9.3	emr-4.9.2	emr-4.9.1	emr-4.8.5
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

Notas de lançamento do 4.9.3

As notas de lançamento a seguir incluem informações para a versão EMR 4.9.3 da Amazon. As alterações são relativas à versão EMR 4.9.2 da Amazon.

Data da versão inicial: 22 de janeiro de 2018

Alterações, melhorias e problemas resolvidos

- Atualizou o kernel Amazon Linux do default Amazon Linux AMI para Amazon para resolver vulnerabilidades associadas EMR à execução especulativa (CVE-2017-5715, -2017-5753 e -2017-5754). CVE CVE Para ter mais informações, consulte <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

4.9.3 versões de componentes

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>emr-ddb</code>	4.3.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.2.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.3.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.4.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.

Componente	Version (Versão)	Descrição
emrfs	2.17.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-2	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-2	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httpfs-server	2.7.3-amzn-2	HTTPendpoint para HDFS operações.

Componente	Version (Versão)	Descrição
hadoop-kms-server	2.7.3-amzn-2	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.2.2	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.2.2	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.2.2	HBasecliente de linha de comando.
hbase-rest-server	1.2.2	Serviço que fornece um RESTful HTTP endpoint paraHBase.

Componente	Version (Versão)	Descrição
hbase-thrift-server	1.2.2	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	1.0.0-amzn-9	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	1.0.0-amzn-9	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	1.0.0-amzn-9	HTTPendpoint fornecendo o uma REST interface paraHCatalog.
hive-client	1.0.0-amzn-9	O cliente da linha de comando do Hive.
hive-metastore-server	1.0.0-amzn-9	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	1.0.0-amzn-9	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.7.1-amzn-7	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.12.2	A biblioteca de Machine Learning.

Componente	Version (Versão)	Descrição
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
oozie-client	4.2.0	O cliente da linha de comando do Oozie.
oozie-server	4.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.7.0- -1,2 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.7.0- -1,2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.157.1	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.157.1	O serviço que executa partes de uma consulta.
pig-client	0.14.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	1.6.3	Os clientes da linha de comando do Spark.
spark-history-server	1.6.3	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.

Componente	Version (Versão)	Descrição
spark-on-yarn	1.6.3	Mecanismo de execução na memória para YARN.
spark-yarn-slave	1.6.3	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.
zeppelin-server	0.6.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.9	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.9	ZooKeeper cliente de linha de comando.

4.9.3 classificações de configuração

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-4.9.3

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.

Classificações	Descrição
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j	Altere os valores nas propriedades ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j	Altere os valores no arquivo hive-exec-log4j.properties do Hive.
hive-log4j	Alterar os valores no arquivo hive-log4j.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.

Classificações	Descrição
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.

Classificações	Descrição
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.

Classificações	Descrição
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.

Classificações	Descrição
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 4.9.2

4.9.2 versões do aplicativo

As seguintes aplicações são compatíveis com esta versão: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) e [ZooKeeper-Sandbox](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-4.9.2	emr-4.9.1	emr-4.8.5	emr-4.8.4
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas

	emr-4.9.2	emr-4.9.1	emr-4.8.5	emr-4.8.4
Scala	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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	-	-	-	-
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

	emr-4.9.2	emr-4.9.1	emr-4.8.5	emr-4.8.4
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

Notas de lançamento do 4.9.2

As notas de lançamento a seguir incluem informações para a versão EMR 4.9.2 da Amazon. As alterações são relativas à versão EMR 4.9.1 da Amazon.

Data do release: 13 de julho de 2017

Pequenas alterações, correções de erros e melhorias foram feitas nesta versão.

4.9.2 versões de componentes

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion-amzn-EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>emr-ddb</code>	4.3.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.2.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.3.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.4.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
<code>emrfs</code>	2.17.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
<code>ganglia-monitor</code>	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop,

Componente	Version (Versão)	Descrição
		juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-2	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-2	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httpfs-server	2.7.3-amzn-2	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-2	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API

Componente	Version (Versão)	Descrição
hadoop-mapred	2.7.3-amzn-2	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.2.2	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.2.2	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.2.2	HBasecliente de linha de comando.
hbase-rest-server	1.2.2	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	1.2.2	Serviço que fornece um endpoint Thrift para. HBase

Componente	Version (Versão)	Descrição
hcatalog-client	1.0.0-amzn-9	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	1.0.0-amzn-9	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	1.0.0-amzn-9	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	1.0.0-amzn-9	O cliente da linha de comando do Hive.
hive-metastore-server	1.0.0-amzn-9	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	1.0.0-amzn-9	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.7.1-amzn-7	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.12.2	A biblioteca de Machine Learning.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.

Componente	Version (Versão)	Descrição
oozie-client	4.2.0	O cliente da linha de comando do Oozie.
oozie-server	4.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.7.0- -1,2 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.7.0- -1,2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.157.1	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.157.1	O serviço que executa partes de uma consulta.
pig-client	0.14.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	1.6.3	Os clientes da linha de comando do Spark.
spark-history-server	1.6.3	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	1.6.3	Mecanismo de execução na memória para YARN.

Componente	Version (Versão)	Descrição
spark-yarn-slave	1.6.3	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.6.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.9	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.9	ZooKeeper cliente de linha de comando.

4.9.2 classificações de configuração

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-4.9.2

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.

Classificações	Descrição
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j	Altere os valores nas propriedades ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j	Altere os valores no arquivo hive-exec-log4j.properties do Hive.
hive-log4j	Alterar os valores no arquivo hive-log4j.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.

Classificações	Descrição
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.

Classificações	Descrição
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.

Classificações	Descrição
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.

Classificações	Descrição
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 4.9.1

4.9.1 versões do aplicativo

As seguintes aplicações são compatíveis com esta versão: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) e [ZooKeeper-Sandbox](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-4.9.1	emr-4.8.5	emr-4.8.4	emr-4.8.3
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas

	emr-4.9.1	emr-4.8.5	emr-4.8.4	emr-4.8.3
Scala	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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	-	-	-	-
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

	emr-4.9.1	emr-4.8.5	emr-4.8.4	emr-4.8.3
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

Notas de lançamento da versão 4.9.1

As notas de lançamento a seguir incluem informações para a versão EMR 4.9.1 da Amazon. As alterações são relativas à versão EMR 4.8.4 da Amazon.

Data do release: 10 de abril de 2017

Problemas conhecidos das versões anteriores que foram resolvidos

- [Portas traseiras de HIVE-9976 e -10106 HIVE](#)
- Corrigido um problema em YARN que um grande número de nós (maior que 2.000) e contêineres (maior que 5.000) causava um out-of-memory erro, por exemplo: "Exception in thread main java.lang.OutOfMemoryError".

Alterações e melhorias

- Os EMR lançamentos da Amazon agora são baseados no Amazon Linux 2017.03. Para ter mais informações, consulte <https://aws.amazon.com/amazon-linux-ami/2017.03-release-notes/>.
- O Python 2.6 foi removido da imagem Linux EMR básica da Amazon. Você pode instalar o Python 2.6 manualmente, se necessário.

4.9.1 versões de componentes

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
emr-ddb	4.2.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.2.0	Bibliotecas convenientes para o ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
emr-kinesis	3.3.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.4.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.15.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-2	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-2	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-2	HDFScliente e biblioteca de linha de comando

Componente	Version (Versão)	Descrição
hadoop-hdfs-namenode	2.7.3-amzn-2	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httfs-server	2.7.3-amzn-2	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-2	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-2	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-2	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-2	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-2	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.2.2	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.2.2	Serviço para atender a uma ou mais HBase regiões.

Componente	Version (Versão)	Descrição
hbase-client	1.2.2	HBase cliente de linha de comando.
hbase-rest-server	1.2.2	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.2.2	Serviço que fornece um endpoint Thrift para HBase.
hcatalog-client	1.0.0-amzn-9	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	1.0.0-amzn-9	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	1.0.0-amzn-9	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	1.0.0-amzn-9	O cliente da linha de comando do Hive.
hive-metastore-server	1.0.0-amzn-9	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	1.0.0-amzn-9	O serviço que aceita as consultas do Hive como solicitações da web.

Componente	Version (Versão)	Descrição
hue-server	3.7.1-amzn-7	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.12.2	A biblioteca de Machine Learning.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
oozie-client	4.2.0	O cliente da linha de comando do Oozie.
oozie-server	4.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.7.0- -1,2 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.7.0- -1,2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.157.1	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.157.1	O serviço que executa partes de uma consulta.
pig-client	0.14.0-amzn-0	O cliente da linha de comando do Pig.

Componente	Version (Versão)	Descrição
spark-client	1.6.3	Os clientes da linha de comando do Spark.
spark-history-server	1.6.3	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	1.6.3	Mecanismo de execução na memória para YARN.
spark-yarn-slave	1.6.3	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.
zeppelin-server	0.6.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.9	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.9	ZooKeeper cliente de linha de comando.

4.9.1 classificações de configuração

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-4.9.1

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase <code>hbase-log4j.properties</code> .
hbase-metrics	Altere os valores no arquivo HBase <code>hadoop-metrics2-hbase.properties</code> do.
hbase-policy	Altere os valores no HBase arquivo <code>hbase-policy.xml</code> do.

Classificações	Descrição
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j	Altere os valores nas propriedades ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j	Altere os valores no arquivo hive-exec-log4j.properties do Hive.
hive-log4j	Alterar os valores no arquivo hive-log4j.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue

Classificações	Descrição
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.

Classificações	Descrição
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.

Classificações	Descrição
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.

Classificações	Descrição
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 4.8.5

4.8.5 versões do aplicativo

As seguintes aplicações são compatíveis com esta versão: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) e [ZooKeeper-Sandbox](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-4.8.5	emr-4.8.4	emr-4.8.3	emr-4.8.2
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75

	emr-4.8.5	emr-4.8.4	emr-4.8.3	emr-4.8.2
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
Scala	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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	-	-	-	-
Mahout	0.12.2	0.12.2	0.12.2	0.12.2
Oozie	-	-	-	-

	emr-4.8.5	emr-4.8.4	emr-4.8.3	emr-4.8.2
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	0.152.3
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

Notas de lançamento do 4.8.5

Esta é uma versão de patch para adicionar a autenticação AWS Signature versão 4 para solicitações ao Amazon S3. Todos os aplicativos e componentes são iguais aos da EMR versão anterior da Amazon.

⚠ Important

Nesta versão de lançamento, a Amazon EMR usa o AWS Signature versão 4 exclusivamente para autenticar solicitações para o Amazon S3. Para obter mais informações, consulte [Novidades](#).

4.8.5 versões de componentes

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>emr-ddb</code>	4.2.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.2.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.2.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.4.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.

Componente	Version (Versão)	Descrição
emrfs	2.14.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-1	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httpfs-server	2.7.3-amzn-1	HTTPendpoint para HDFS operações.

Componente	Version (Versão)	Descrição
hadoop-kms-server	2.7.3-amzn-1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.2.2	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.2.2	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.2.2	HBasecliente de linha de comando.
hbase-rest-server	1.2.2	Serviço que fornece um RESTful HTTP endpoint paraHBase.

Componente	Version (Versão)	Descrição
hbase-thrift-server	1.2.2	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	1.0.0-amzn-8	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	1.0.0-amzn-8	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	1.0.0-amzn-8	HTTPendpoint fornecendo o uma REST interface paraHCatalog.
hive-client	1.0.0-amzn-8	O cliente da linha de comando do Hive.
hive-metastore-server	1.0.0-amzn-8	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	1.0.0-amzn-8	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.7.1-amzn-7	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.12.2	A biblioteca de Machine Learning.

Componente	Version (Versão)	Descrição
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.
oozie-client	4.2.0	O cliente da linha de comando do Oozie.
oozie-server	4.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.7.0- -1,2 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.7.0- -1,2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.157.1	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.157.1	O serviço que executa partes de uma consulta.
pig-client	0.14.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	1.6.3	Os clientes da linha de comando do Spark.
spark-history-server	1.6.3	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.

Componente	Version (Versão)	Descrição
spark-on-yarn	1.6.3	Mecanismo de execução na memória para YARN.
spark-yarn-slave	1.6.3	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTP Servidor Apache.
zeppelin-server	0.6.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.9	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.9	ZooKeeper cliente de linha de comando.

4.8.5 classificações de configuração

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-4.8.5

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.

Classificações	Descrição
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j	Altere os valores nas propriedades ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j	Altere os valores no arquivo hive-exec-log4j.properties do Hive.
hive-log4j	Alterar os valores no arquivo hive-log4j.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.

Classificações	Descrição
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.

Classificações	Descrição
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.

Classificações	Descrição
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.

Classificações	Descrição
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 4.8.4

4.8.4 versões do aplicativo

As seguintes aplicações são compatíveis com esta versão: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) e [ZooKeeper-Sandbox](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-4.8.4	emr-4.8.3	emr-4.8.2	emr-4.8.1
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas

	emr-4.8.4	emr-4.8.3	emr-4.8.2	emr-4.8.1
Scala	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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	-	-	-	-
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

	emr-4.8.4	emr-4.8.3	emr-4.8.2	emr-4.8.1
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

Notas de lançamento do 4.8.4

As notas de lançamento a seguir incluem informações para a versão EMR 4.8.4 da Amazon. As alterações são relativas à versão EMR 4.8.3 da Amazon.

Data da versão: 7 de fevereiro de 2017

Pequenas alterações, correções de erros e melhorias foram feitas nesta versão.

4.8.4 versões de componentes

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion-amzn-EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>emr-ddb</code>	4.2.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.2.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.2.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.4.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
<code>emrfs</code>	2.14.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
<code>ganglia-monitor</code>	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop,

Componente	Version (Versão)	Descrição
		juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-1	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httpfs-server	2.7.3-amzn-1	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API

Componente	Version (Versão)	Descrição
hadoop-mapred	2.7.3-amzn-1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.2.2	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.2.2	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.2.2	HBasecliente de linha de comando.
hbase-rest-server	1.2.2	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	1.2.2	Serviço que fornece um endpoint Thrift para. HBase

Componente	Version (Versão)	Descrição
hcatalog-client	1.0.0-amzn-8	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	1.0.0-amzn-8	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	1.0.0-amzn-8	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	1.0.0-amzn-8	O cliente da linha de comando do Hive.
hive-metastore-server	1.0.0-amzn-8	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	1.0.0-amzn-8	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.7.1-amzn-7	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.12.2	A biblioteca de Machine Learning.
mysql-server	5.5.54+	Meu servidor SQL de banco de dados.

Componente	Version (Versão)	Descrição
oozie-client	4.2.0	O cliente da linha de comando do Oozie.
oozie-server	4.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.7.0- -1,2 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.7.0- -1,2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.157.1	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.157.1	O serviço que executa partes de uma consulta.
pig-client	0.14.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	1.6.3	Os clientes da linha de comando do Spark.
spark-history-server	1.6.3	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	1.6.3	Mecanismo de execução na memória para YARN.

Componente	Version (Versão)	Descrição
spark-yarn-slave	1.6.3	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.25+	HTTPServidor Apache.
zeppelin-server	0.6.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.9	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.9	ZooKeeper cliente de linha de comando.

4.8.4 classificações de configuração

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-4.8.4

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.

Classificações	Descrição
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j	Altere os valores nas propriedades ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j	Altere os valores no arquivo hive-exec-log4j.properties do Hive.
hive-log4j	Alterar os valores no arquivo hive-log4j.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.

Classificações	Descrição
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.

Classificações	Descrição
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.

Classificações	Descrição
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.

Classificações	Descrição
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 4.8.3

4.8.3 versões do aplicativo

As seguintes aplicações são compatíveis com esta versão: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) e [ZooKeeper-Sandbox](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-4.8.3	emr-4.8.2	emr-4.8.1	emr-4.8.0
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas

	emr-4.8.3	emr-4.8.2	emr-4.8.1	emr-4.8.0
Scala	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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	-	-	-	-
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

	emr-4.8.3	emr-4.8.2	emr-4.8.1	emr-4.8.0
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
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.9	3.4.8	3.4.8	3.4.8

Notas de lançamento do 4.8.3

As notas de lançamento a seguir incluem informações para a versão EMR 4.8.3 da Amazon. As alterações são relativas à versão EMR 4.8.2 da Amazon.

Data do release: 29 de dezembro de 2016

Atualizações

- Atualizado para Presto 0.157.1. Para obter mais informações, consulte [Notas de release do Presto](#) na documentação do Presto.
- Atualizado para Spark 1.6.3. Para obter mais informações, consulte as [Notas de release do Spark](#) na documentação do Apache Spark.
- Atualizado para ZooKeeper 3.4.9. Para obter mais informações, consulte as [notas de ZooKeeper lançamento](#) na ZooKeeper documentação do Apache.

Alterações e melhorias

- Foi adicionado suporte para o tipo de EC2 instância Amazon m4.16xlarge na Amazon EMR versão 4.8.3 e posterior, excluindo 5.0.0, 5.0.3 e 5.2.0.
- Os EMR lançamentos da Amazon agora são baseados no Amazon Linux 2016.09. Para ter mais informações, consulte <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.

Problemas conhecidos das versões anteriores que foram resolvidos

- Corrigido um problema no Hadoop em que o ReplicationMonitor thread podia ficar preso por um longo tempo devido a uma corrida entre a replicação e a exclusão do mesmo arquivo em um grande cluster.
- Corrigido um problema em que ControlledJob # toString falhava com uma exceção de ponteiro nulo (NPE) quando o status do trabalho não era atualizado com êxito.

4.8.3 versões de componentes

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto

nomeado myapp-component com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como 2.2-amzn-2.

Componente	Version (Versão)	Descrição
emr-ddb	4.2.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.2.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.2.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.4.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.13.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.

Componente	Version (Versão)	Descrição
hadoop-client	2.7.3-amzn-1	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-https-server	2.7.3-amzn-1	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.3-amzn-1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-1	Serviço para recuperar informações atuais e históricas para YARN aplicativos.

Componente	Version (Versão)	Descrição
hbase-hmaster	1.2.2	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.2.2	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.2.2	HBase cliente de linha de comando.
hbase-rest-server	1.2.2	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.2.2	Serviço que fornece um endpoint Thrift para HBase.
hcatalog-client	1.0.0-amzn-8	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	1.0.0-amzn-8	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	1.0.0-amzn-8	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	1.0.0-amzn-8	O cliente da linha de comando do Hive.

Componente	Version (Versão)	Descrição
hive-metastore-server	1.0.0-amzn-8	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	1.0.0-amzn-8	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.7.1-amzn-7	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.12.2	A biblioteca de Machine Learning.
mysql-server	5.5.52	Meu servidor SQL de banco de dados.
oozie-client	4.2.0	O cliente da linha de comando do Oozie.
oozie-server	4.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.7.0- -1,2 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.7.0- -1,2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API

Componente	Version (Versão)	Descrição
presto-coordinator	0.157.1	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.157.1	O serviço que executa partes de uma consulta.
pig-client	0.14.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	1.6.3	Os clientes da linha de comando do Spark.
spark-history-server	1.6.3	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	1.6.3	Mecanismo de execução na memória para YARN.
spark-yarn-slave	1.6.3	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.23	HTTP Servidor Apache.

Componente	Version (Versão)	Descrição
zeppelin-server	0.6.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.9	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.9	ZooKeeper cliente de linha de comando.

4.8.3 classificações de configuração

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-4.8.3

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Hadoop.

Classificações	Descrição
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j	Altere os valores nas propriedades ebHCat log4j.properties de HCatalog W.

Classificações	Descrição
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j	Altere os valores no arquivo hive-exec-log4j.properties do Hive.
hive-log4j	Alterar os valores no arquivo hive-log4j.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.

Classificações	Descrição
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.

Classificações	Descrição
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.

Classificações	Descrição
spark-env	Alterar os valores no ambiente do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 4.8.2

4.8.2 versões do aplicativo

As seguintes aplicações são compatíveis com esta versão: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) e [ZooKeeper-Sandbox](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-4.8.2	emr-4.8.1	emr-4.8.0	emr-4.7.4
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
Scala	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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

	emr-4.8.2	emr-4.8.1	emr-4.8.0	emr-4.7.4
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	-	-	-	-
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

	emr-4.8.2	emr-4.8.1	emr-4.8.0	emr-4.7.4
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

Notas de lançamento do 4.8.2

As notas de lançamento a seguir incluem informações para a versão EMR 4.8.2 da Amazon. As alterações são relativas à versão EMR 4.8.0 da Amazon.

Data do release: 24 de outubro de 2016

Atualizações

- Atualizado para Hadoop 2.7.3
- Atualizado para Presto 0.152.3, que inclui o suporte para a interface da web do Presto. Você pode acessar a interface da web do Presto no coordenador do Presto usando a porta 8889. Para obter mais informações sobre a interface da web do Presto, consulte [Interface da web](#) na documentação do Presto.
- Os EMR lançamentos da Amazon agora são baseados no Amazon Linux 2016.09. Para ter mais informações, consulte <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.

4.8.2 versões de componentes

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion-amzn-EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>emr-ddb</code>	4.1.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.1.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.2.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.4.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
<code>emrfs</code>	2.10.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
<code>ganglia-monitor</code>	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop,

Componente	Version (Versão)	Descrição
		juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.3-amzn-0	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.3-amzn-0	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.3-amzn-0	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.3-amzn-0	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-https-server	2.7.3-amzn-0	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.3-amzn-0	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API

Componente	Version (Versão)	Descrição
hadoop-mapred	2.7.3-amzn-0	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.3-amzn-0	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.3-amzn-0	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.3-amzn-0	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.2.2	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.2.2	Serviço para atender a uma ou mais HBase regiões.
hbase-client	1.2.2	HBasecliente de linha de comando.
hbase-rest-server	1.2.2	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	1.2.2	Serviço que fornece um endpoint Thrift para. HBase

Componente	Version (Versão)	Descrição
hcatalog-client	1.0.0-amzn-7	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	1.0.0-amzn-7	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	1.0.0-amzn-7	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	1.0.0-amzn-7	O cliente da linha de comando do Hive.
hive-metastore-server	1.0.0-amzn-7	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	1.0.0-amzn-7	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.7.1-amzn-7	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.12.2	A biblioteca de Machine Learning.
mysql-server	5.5.52	Meu servidor SQL de banco de dados.

Componente	Version (Versão)	Descrição
oozie-client	4.2.0	O cliente da linha de comando do Oozie.
oozie-server	4.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.7.0- -1,2 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.7.0- -1,2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.152.3	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.152.3	O serviço que executa partes de uma consulta.
pig-client	0.14.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	1.6.2	Os clientes da linha de comando do Spark.
spark-history-server	1.6.2	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	1.6.2	Mecanismo de execução na memória para YARN.

Componente	Version (Versão)	Descrição
spark-yarn-slave	1.6.2	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.23	HTTPServidor Apache.
zeppelin-server	0.6.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.8	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.8	ZooKeeper cliente de linha de comando.

4.8.2 classificações de configuração

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-4.8.2

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.

Classificações	Descrição
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j	Altere os valores nas propriedades ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j	Altere os valores no arquivo hive-exec-log4j.properties do Hive.
hive-log4j	Alterar os valores no arquivo hive-log4j.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.

Classificações	Descrição
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.

Classificações	Descrição
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.

Classificações	Descrição
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.

Classificações	Descrição
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 4.8.1

4.8.1 versões do aplicativo

As seguintes aplicações são compatíveis com esta versão: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) e [ZooKeeper-Sandbox](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-4.8.1	emr-4.8.0	emr-4.7.4	emr-4.7.3
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas

	emr-4.8.1	emr-4.8.0	emr-4.7.4	emr-4.7.3
Scala	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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	-	-	-	-
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

	emr-4.8.1	emr-4.8.0	emr-4.7.4	emr-4.7.3
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)	-	-	-	-
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

Notas de lançamento do 4.8.1

4.8.1 versões de componentes

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou

aws. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado myapp-component com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como 2.2-amzn-2.

Componente	Version (Versão)	Descrição
emr-ddb	3.2.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.1.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.2.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.4.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.9.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas

Componente	Version (Versão)	Descrição
		de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.2-amzn-4	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.2-amzn-4	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.2-amzn-4	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.2-amzn-4	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httpfs-server	2.7.2-amzn-4	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.2-amzn-4	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.2-amzn-4	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.2-amzn-4	YARNserviço para gerenciar contêineres em um nó individual.

Componente	Version (Versão)	Descrição
hadoop-yarn-resourcemanager	2.7.2-amzn-4	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.2-amzn-4	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.2.2	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.2.2	Serviço para atender uma ou mais HBase regiões.
hbase-client	1.2.2	HBasecliente de linha de comando.
hbase-rest-server	1.2.2	Serviço que fornece um RESTful HTTP endpoint paraHBase.
hbase-thrift-server	1.2.2	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	1.0.0-amzn-7	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	1.0.0-amzn-7	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.

Componente	Version (Versão)	Descrição
hcatalog-webhcat-server	1.0.0-amzn-7	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	1.0.0-amzn-7	O cliente da linha de comando do Hive.
hive-metastore-server	1.0.0-amzn-7	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	1.0.0-amzn-7	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.7.1-amzn-7	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.12.2	A biblioteca de Machine Learning.
mysql-server	5.5.51	Meu servidor SQL de banco de dados.
oozie-client	4.2.0	O cliente da linha de comando do Oozie.
oozie-server	4.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.7.0- -1,2 HBase	As bibliotecas do phoenix para servidor e cliente

Componente	Version (Versão)	Descrição
phoenix-query-server	4.7.0- -1,2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.151	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.151	O serviço que executa partes de uma consulta.
pig-client	0.14.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	1.6.2	Os clientes da linha de comando do Spark.
spark-history-server	1.6.2	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	1.6.2	Mecanismo de execução na memória para YARN.
spark-yarn-slave	1.6.2	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.

Componente	Version (Versão)	Descrição
webserver	2.4.23	HTTPServidor Apache.
zeppelin-server	0.6.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.8	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.8	ZooKeeper cliente de linha de comando.

4.8.1 classificações de configuração

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

classificações do emr-4.8.1

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.

Classificações	Descrição
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.

Classificações	Descrição
hcatalog-webhcat-log4j	Altere os valores nas propriedades ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j	Altere os valores no arquivo hive-exec-log4j.properties do Hive.
hive-log4j	Alterar os valores no arquivo hive-log4j.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.

Classificações	Descrição
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.

Classificações	Descrição
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.

Classificações	Descrição
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 4.8.0

Versões do aplicativo 4.8.0

As seguintes aplicações são compatíveis com esta versão: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) e [ZooKeeper-Sandbox](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-4.8.0	emr-4.7.4	emr-4.7.3	emr-4.7.2
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
Scala	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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

	emr-4.8.0	emr-4.7.4	emr-4.7.3	emr-4.7.2
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	0.12.2	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

	emr-4.8.0	emr-4.7.4	emr-4.7.3	emr-4.7.2
TensorFlow	-	-	-	-
Tez	0.8.4	0.8.3	0.8.3	0.8.3
Trino (PrestoSQL)	-	-	-	-
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

Notas de lançamento do 4.8.0

As notas de lançamento a seguir incluem informações para a versão EMR 4.8.0 da Amazon. As alterações são relativas à EMR versão 4.7.2 da Amazon.

Data do release: 7 de setembro de 2016

Atualizações

- Atualizado para 1.2.2 HBase
- Atualizado para Presto-Sandbox 0.151
- Atualizado para Tez 0.8.4
- Atualizado para Zeppelin-Sandbox 0.6.1

Alterações e melhorias

- Corrigido um problema YARN em que eles ApplicationMaster tentavam limpar contêineres que não existem mais porque suas instâncias foram encerradas.
- Foram corrigidas as ações hive-server2 URL para Hive2 nos exemplos do Oozie.

- Adicionado o suporte para catálogos Presto adicionais.
- [Patches retroportados: HIVE-8948, HIVE -12679, -13405, -3116, -12689 HIVE PHOENIX HADOOP](#)
- Adicionado o suporte para configurações de segurança, que permitem criar e aplicar opções de criptografia com mais facilidade. Para obter mais informações, consulte [Criptografia de dados](#).

Versões de componentes 4.8.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>emr-ddb</code>	3.2.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.1.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.2.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.4.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.

Componente	Version (Versão)	Descrição
emrfs	2.9.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.2-amzn-4	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.2-amzn-4	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.2-amzn-4	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.2-amzn-4	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httpfs-server	2.7.2-amzn-4	HTTPendpoint para HDFS operações.

Componente	Version (Versão)	Descrição
hadoop-kms-server	2.7.2-amzn-4	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.2-amzn-4	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.2-amzn-4	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.2-amzn-4	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.2-amzn-4	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.2.2	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.2.2	Serviço para atender uma ou mais HBase regiões.
hbase-client	1.2.2	HBasecliente de linha de comando.
hbase-rest-server	1.2.2	Serviço que fornece um RESTful HTTP endpoint paraHBase.

Componente	Version (Versão)	Descrição
hbase-thrift-server	1.2.2	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	1.0.0-amzn-7	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	1.0.0-amzn-7	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	1.0.0-amzn-7	HTTPendpoint fornecendo o uma REST interface paraHCatalog.
hive-client	1.0.0-amzn-7	O cliente da linha de comando do Hive.
hive-metastore-server	1.0.0-amzn-7	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	1.0.0-amzn-7	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.7.1-amzn-7	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.12.2	A biblioteca de Machine Learning.

Componente	Version (Versão)	Descrição
mysql-server	5.5.51	Meu servidor SQL de banco de dados.
oozie-client	4.2.0	O cliente da linha de comando do Oozie.
oozie-server	4.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.7.0- -1,2 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.7.0- -1,2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.151	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.151	O serviço que executa partes de uma consulta.
pig-client	0.14.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	1.6.2	Os clientes da linha de comando do Spark.
spark-history-server	1.6.2	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.

Componente	Version (Versão)	Descrição
spark-on-yarn	1.6.2	Mecanismo de execução na memória para YARN.
spark-yarn-slave	1.6.2	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.4	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.23	HTTP Servidor Apache.
zeppelin-server	0.6.1	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.8	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.8	ZooKeeper cliente de linha de comando.

4.8.0 classificações de configuração

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-4.8.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.

Classificações	Descrição
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j	Altere os valores nas propriedades ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j	Altere os valores no arquivo hive-exec-log4j.properties do Hive.
hive-log4j	Alterar os valores no arquivo hive-log4j.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hiveserver2-site	Alterar os valores no arquivo hiveserver2-site.xml do Hive Server2.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.

Classificações	Descrição
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.

Classificações	Descrição
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-blackhole	Alterar os valores no arquivo blackhole.properties do Presto.
presto-connector-cassandra	Alterar os valores no arquivo cassandra.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
presto-connector-jmx	Alterar os valores no arquivo jmx.properties do Presto.
presto-connector-kafka	Alterar os valores no arquivo kafka.properties do Presto.
presto-connector-localfile	Alterar os valores no arquivo localfile.properties do Presto.
presto-connector-mongodb	Alterar os valores no arquivo mongodb.properties do Presto.
presto-connector-mysql	Alterar os valores no arquivo mysql.properties do Presto.
presto-connector-postgresql	Alterar os valores no arquivo postgresql.properties do Presto.

Classificações	Descrição
presto-connector-raptor	Alterar os valores no arquivo raptor.properties do Presto.
presto-connector-redis	Alterar os valores no arquivo redis.properties do Presto.
presto-connector-tpch	Alterar os valores no arquivo tpch.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.

Classificações	Descrição
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 4.7.4

4.7.4 versões do aplicativo

As seguintes aplicações são compatíveis com esta versão: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) e [ZooKeeper-Sandbox](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-4.7.4	emr-4.7.3	emr-4.7.2	emr-4.7.1
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas

	emr-4.7.4	emr-4.7.3	emr-4.7.2	emr-4.7.1
Scala	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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	-	-	-	-
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

	emr-4.7.4	emr-4.7.3	emr-4.7.2	emr-4.7.1
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

Notas de lançamento do 4.7.4

Esta é uma versão de patch para adicionar a autenticação AWS Signature versão 4 para solicitações ao Amazon S3. Todos os aplicativos e componentes são iguais aos da EMR versão anterior da Amazon.

⚠ Important

Nesta versão de lançamento, a Amazon EMR usa o AWS Signature versão 4 exclusivamente para autenticar solicitações para o Amazon S3. Para obter mais informações, consulte [Novidades](#).

4.7.4 versões de componentes

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>emr-ddb</code>	3.2.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.1.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.2.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.4.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.

Componente	Version (Versão)	Descrição
emrfs	2.8.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.2-amzn-3	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.2-amzn-3	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.2-amzn-3	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.2-amzn-3	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httpfs-server	2.7.2-amzn-3	HTTPendpoint para HDFS operações.

Componente	Version (Versão)	Descrição
hadoop-kms-server	2.7.2-amzn-3	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.2-amzn-3	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.2-amzn-3	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.2-amzn-3	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.2-amzn-3	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.2.1	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.2.1	Serviço para atender uma ou mais HBase regiões.
hbase-client	1.2.1	HBasecliente de linha de comando.
hbase-rest-server	1.2.1	Serviço que fornece um RESTful HTTP endpoint paraHBase.

Componente	Version (Versão)	Descrição
hbase-thrift-server	1.2.1	Serviço que fornece um endpoint Thrift para. HBase
hcatalog-client	1.0.0-amzn-6	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	1.0.0-amzn-6	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	1.0.0-amzn-6	HTTPendpoint fornecendo o uma REST interface paraHCatalog.
hive-client	1.0.0-amzn-6	O cliente da linha de comando do Hive.
hive-metastore-server	1.0.0-amzn-6	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	1.0.0-amzn-6	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.7.1-amzn-7	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.12.2	A biblioteca de Machine Learning.

Componente	Version (Versão)	Descrição
mysql-server	5.5.46	Meu servidor SQL de banco de dados.
oozie-client	4.2.0	O cliente da linha de comando do Oozie.
oozie-server	4.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.7.0- -1,2 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.7.0- -1,2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.148	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.148	O serviço que executa partes de uma consulta.
pig-client	0.14.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	1.6.2	Os clientes da linha de comando do Spark.
spark-history-server	1.6.2	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.

Componente	Version (Versão)	Descrição
spark-on-yarn	1.6.2	Mecanismo de execução na memória para YARN.
spark-yarn-slave	1.6.2	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.3	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.23	HTTP Servidor Apache.
zeppelin-server	0.5.6-incubating	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.8	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.8	ZooKeeper cliente de linha de comando.

4.7.4 classificações de configuração

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-4.7.4

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo capacity-scheduler.xml do Hadoop.
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.

Classificações	Descrição
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j	Altere os valores nas propriedades ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j	Altere os valores no arquivo hive-exec-log4j.properties do Hive.
hive-log4j	Alterar os valores no arquivo hive-log4j.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.

Classificações	Descrição
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.

Classificações	Descrição
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.

Classificações	Descrição
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 4.7.3

4.7.3 versões do aplicativo

As seguintes aplicações são compatíveis com esta versão: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) e [ZooKeeper-Sandbox](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-4.7.3	emr-4.7.2	emr-4.7.1	emr-4.7.0
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.75
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas

	emr-4.7.3	emr-4.7.2	emr-4.7.1	emr-4.7.0
Scala	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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	-	-	-	-
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

	emr-4.7.3	emr-4.7.2	emr-4.7.1	emr-4.7.0
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
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

Notas de lançamento do 4.7.3

4.7.3 versões de componentes

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou

aws. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado myapp-component com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como 2.2-amzn-2.

Componente	Version (Versão)	Descrição
emr-ddb	3.2.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.1.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.2.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.4.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.8.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas

Componente	Version (Versão)	Descrição
		de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.2-amzn-3	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.2-amzn-3	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.2-amzn-3	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.2-amzn-3	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httpfs-server	2.7.2-amzn-3	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.2-amzn-3	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.2-amzn-3	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.2-amzn-3	YARNserviço para gerenciar contêineres em um nó individual.

Componente	Version (Versão)	Descrição
hadoop-yarn-resourcemanager	2.7.2-amzn-3	YARN serviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.2-amzn-3	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.2.1	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.2.1	Serviço para atender uma ou mais HBase regiões.
hbase-client	1.2.1	HBase cliente de linha de comando.
hbase-rest-server	1.2.1	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.2.1	Serviço que fornece um endpoint Thrift para HBase.
hcatalog-client	1.0.0-amzn-6	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	1.0.0-amzn-6	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.

Componente	Version (Versão)	Descrição
hcatalog-webhcat-server	1.0.0-amzn-6	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	1.0.0-amzn-6	O cliente da linha de comando do Hive.
hive-metastore-server	1.0.0-amzn-6	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	1.0.0-amzn-6	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.7.1-amzn-7	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.12.2	A biblioteca de Machine Learning.
mysql-server	5.5.46	Meu servidor SQL de banco de dados.
oozie-client	4.2.0	O cliente da linha de comando do Oozie.
oozie-server	4.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.7.0- -1,2 HBase	As bibliotecas do phoenix para servidor e cliente

Componente	Version (Versão)	Descrição
phoenix-query-server	4.7.0- -1,2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.148	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.148	O serviço que executa partes de uma consulta.
pig-client	0.14.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	1.6.2	Os clientes da linha de comando do Spark.
spark-history-server	1.6.2	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	1.6.2	Mecanismo de execução na memória para YARN.
spark-yarn-slave	1.6.2	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.3	O YARN aplicativo e as bibliotecas tez.

Componente	Version (Versão)	Descrição
webserver	2.4.23	HTTPServidor Apache.
zeppelin-server	0.5.6-incubating	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.8	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.8	ZooKeeper cliente de linha de comando.

4.7.3 classificações de configuração

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

classificações do emr-4.7.3

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.

Classificações	Descrição
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.

Classificações	Descrição
hcatalog-webhcat-log4j	Altere os valores nas propriedades ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j	Altere os valores no arquivo hive-exec-log4j.properties do Hive.
hive-log4j	Alterar os valores no arquivo hive-log4j.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.

Classificações	Descrição
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.

Classificações	Descrição
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 4.7.2

4.7.2 versões do aplicativo

As seguintes aplicações são compatíveis com esta versão: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) e [ZooKeeper-Sandbox](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-4.7.2	emr-4.7.1	emr-4.7.0	emr-4.6.1
AWS SDK para Java	1.10.75	1.10.75	1.10.75	1.10.27
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
Scala	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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

	emr-4.7.2	emr-4.7.1	emr-4.7.0	emr-4.6.1
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	-	-	-	-
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	-
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

	emr-4.7.2	emr-4.7.1	emr-4.7.0	emr-4.6.1
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

Notas de lançamento do 4.7.2

As notas de lançamento a seguir incluem informações sobre o Amazon EMR 4.7.2.

Data do release: 15 de julho de 2016

Recursos

- Atualizado para Mahout 0.12.2
- Atualizado para Presto 0.148
- Atualizado para Spark 1.6.2
- Agora você pode criar um `AWSCredentialsProvider` para uso com EMRFS o uso de um URI como parâmetro. Para obter mais informações, consulte [Criar um AWSCredentialsProvider formulário EMRFS](#).
- EMRFS agora permite que os usuários configurem um endpoint personalizado do DynamoDB para seus metadados do Consistent View usando a propriedade em `fs.s3.consistent.dynamodb.endpoint emrfs-site.xml`
- Adicionado um script em `/usr/bin` chamado `spark-example`, que encapsula `/usr/lib/spark/spark/bin/run-example` para que você possa executar exemplos diretamente. Por

exemplo, para executar o SparkPi exemplo que vem com a distribuição do Spark, você pode executar a `spark-example SparkPi 100` partir da linha de comando ou usando `command-runner.jar` como uma etapa noAPI.

Problemas conhecidos das versões anteriores que foram resolvidos

- Corrigido um problema em que o Oozie não tinha o `spark-assembly.jar` no local correto quando o Spark também estava instalado, o que resultava em falha para iniciar aplicativos do Spark com o Oozie.
- Corrigido um problema com o registro baseado no Spark Log4J em contêineres. YARN

4.7.2 versões de componentes

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>emr-ddb</code>	3.2.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.1.0	Bibliotecas convenientes para o ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
emr-kinesis	3.2.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.4.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.8.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.2-amzn-3	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.2-amzn-3	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.2-amzn-3	HDFScliente e biblioteca de linha de comando

Componente	Version (Versão)	Descrição
hadoop-hdfs-namenode	2.7.2-amzn-3	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httfs-server	2.7.2-amzn-3	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.2-amzn-3	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.2-amzn-3	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.2-amzn-3	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.2-amzn-3	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.2-amzn-3	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.2.1	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.2.1	Serviço para atender uma ou mais HBase regiões.

Componente	Version (Versão)	Descrição
hbase-client	1.2.1	HBase cliente de linha de comando.
hbase-rest-server	1.2.1	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.2.1	Serviço que fornece um endpoint Thrift para HBase.
hcatalog-client	1.0.0-amzn-6	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	1.0.0-amzn-6	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	1.0.0-amzn-6	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	1.0.0-amzn-6	O cliente da linha de comando do Hive.
hive-metastore-server	1.0.0-amzn-6	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	1.0.0-amzn-6	O serviço que aceita as consultas do Hive como solicitações da web.

Componente	Version (Versão)	Descrição
hue-server	3.7.1-amzn-7	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.12.2	A biblioteca de Machine Learning.
mysql-server	5.5.46	Meu servidor SQL de banco de dados.
oozie-client	4.2.0	O cliente da linha de comando do Oozie.
oozie-server	4.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.7.0- -1,2 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.7.0- -1,2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.148	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.148	O serviço que executa partes de uma consulta.
pig-client	0.14.0-amzn-0	O cliente da linha de comando do Pig.

Componente	Version (Versão)	Descrição
spark-client	1.6.2	Os clientes da linha de comando do Spark.
spark-history-server	1.6.2	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	1.6.2	Mecanismo de execução na memória para YARN.
spark-yarn-slave	1.6.2	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.3	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.23	HTTP Servidor Apache.
zeppelin-server	0.5.6-incubating	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.8	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.8	ZooKeeper cliente de linha de comando.

4.7.2 classificações de configuração

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-4.7.2

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Hadoop.
hadoop-ssl-server	Alterar a configuração do servidor SSL no Hadoop
hadoop-ssl-client	Alterar a configuração do cliente SSL no Hadoop
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase <code>hbase-log4j.properties</code> .
hbase-metrics	Altere os valores no arquivo HBase <code>hadoop-metrics2-hbase.properties</code> do.
hbase-policy	Altere os valores no HBase arquivo <code>hbase-policy.xml</code> do.

Classificações	Descrição
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j	Altere os valores nas propriedades ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j	Altere os valores no arquivo hive-exec-log4j.properties do Hive.
hive-log4j	Alterar os valores no arquivo hive-log4j.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.

Classificações	Descrição
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.

Classificações	Descrição
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.

Classificações	Descrição
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 4.7.1

4.7.1 versões do aplicativo

As seguintes aplicações são compatíveis com esta versão: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) e [ZooKeeper-Sandbox](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-4.7.1	emr-4.7.0	emr-4.6.1	emr-4.6.0
AWS SDK para Java	1.10.75	1.10.75	1.10.27	1.10.27
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
Scala	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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	-	-	-	-

	emr-4.7.1	emr-4.7.0	emr-4.6.1	emr-4.6.0
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	-	-
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

Notas de lançamento do 4.7.1

As notas de lançamento a seguir incluem informações sobre a Amazon EMR 4.7.1.

Data do release: 10 de junho de 2016

Problemas conhecidos das versões anteriores que foram resolvidos

- Corrigido um problema que estendia o tempo de inicialização dos clusters lançados em um VPC com sub-redes privadas. O bug afetou apenas os clusters lançados com a EMR versão 4.7.0 da Amazon.
- Corrigido um problema que manipulava indevidamente a listagem de arquivos na Amazon EMR para clusters lançados com a versão EMR 4.7.0 da Amazon.

4.7.1 versões de componentes

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
emr-ddb	3.1.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.0.0	Bibliotecas convenientes para o ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
emr-kinesis	3.2.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.4.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.7.1	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.2-amzn-2	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.2-amzn-2	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.2-amzn-2	HDFScliente e biblioteca de linha de comando

Componente	Version (Versão)	Descrição
hadoop-hdfs-namenode	2.7.2-amzn-2	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httfs-server	2.7.2-amzn-2	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.2-amzn-2	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.2-amzn-2	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.2-amzn-2	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.2-amzn-2	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.2-amzn-2	Serviço para recuperar informações atuais e históricas para YARN aplicativos.
hbase-hmaster	1.2.1	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.2.1	Serviço para atender uma ou mais HBase regiões.

Componente	Version (Versão)	Descrição
hbase-client	1.2.1	HBase cliente de linha de comando.
hbase-rest-server	1.2.1	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.2.1	Serviço que fornece um endpoint Thrift para HBase.
hcatalog-client	1.0.0-amzn-5	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	1.0.0-amzn-5	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	1.0.0-amzn-5	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	1.0.0-amzn-5	O cliente da linha de comando do Hive.
hive-metastore-server	1.0.0-amzn-5	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	1.0.0-amzn-5	O serviço que aceita as consultas do Hive como solicitações da web.

Componente	Version (Versão)	Descrição
hue-server	3.7.1-amzn-7	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.12.0	A biblioteca de Machine Learning.
mysql-server	5.5.46	Meu servidor SQL de banco de dados.
oozie-client	4.2.0	O cliente da linha de comando do Oozie.
oozie-server	4.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.7.0- -1,2 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.7.0- -1,2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API
presto-coordinator	0.147	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.147	O serviço que executa partes de uma consulta.
pig-client	0.14.0-amzn-0	O cliente da linha de comando do Pig.

Componente	Version (Versão)	Descrição
spark-client	1.6.1	Os clientes da linha de comando do Spark.
spark-history-server	1.6.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	1.6.1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	1.6.1	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.3	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.18	HTTP Servidor Apache.
zeppelin-server	0.5.6-incubating	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.8	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.8	ZooKeeper cliente de linha de comando.

4.7.1 classificações de configuração

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-4.7.1

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Hadoop.
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase <code>hbase-log4j.properties</code> .
hbase-metrics	Altere os valores no arquivo HBase <code>hadoop-metrics2-hbase.properties</code> do.
hbase-policy	Altere os valores no HBase arquivo <code>hbase-policy.xml</code> do.
hbase-site	Altere os valores no HBase arquivo <code>hbase-site.xml</code> do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's <code>hdfs-site.xml</code> .

Classificações	Descrição
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j	Altere os valores nas propriedades ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j	Altere os valores no arquivo hive-exec-log4j.properties do Hive.
hive-log4j	Alterar os valores no arquivo hive-log4j.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.

Classificações	Descrição
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.

Classificações	Descrição
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.

Classificações	Descrição
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 4.7.0

Versões do aplicativo 4.7.0

As seguintes aplicações são compatíveis com esta versão: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Phoenix](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Tez](#), [Zeppelin-Sandbox](#) e [ZooKeeper-Sandbox](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-4.7.0	emr-4.6.1	emr-4.6.0	emr-4.5.0
AWS SDK para Java	1.10.75	1.10.27	1.10.27	1.10.27
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas

	emr-4.7.0	emr-4.6.1	emr-4.6.0	emr-4.5.0
Scala	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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	-	-	-	-
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

	emr-4.7.0	emr-4.6.1	emr-4.6.0	emr-4.5.0
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	-

Notas de lançamento do 4.7.0

Important

O Amazon EMR 4.7.0 está obsoleto. Em vez disso, use o Amazon EMR 4.7.1 ou posterior.

Data do release: 2 de junho de 2016

Recursos

- Adicionado o Apache Phoenix 4.7.0
- Adicionado o Apache Tez 0.8.3
- Atualizado para 1.2.1 HBase
- Atualizado para Mahout 0.12.0
- Atualizado para Presto 0.147
- Atualizou o AWS SDK for Java para 1.10.75
- O sinalizador final foi removido da propriedade `mapreduce.cluster.local.dir` em `mapred-site.xml` para permitir que os usuários executem o Pig no modo local.
- JDBCDrivers do Amazon Redshift disponíveis no cluster

JDBCOs drivers do Amazon Redshift agora estão incluídos em `/usr/share/aws/redshift/jdbc`. `/usr/share/aws/redshift/jdbc/RedshiftJDBC41.jar` é o driver do Amazon Redshift JDBC compatível com 4.1. `/usr/share/aws/redshift/jdbc/RedshiftJDBC4.jar` é o driver do Amazon Redshift JDBC compatível com 4.0. Para obter mais informações, consulte [Configurar uma JDBC conexão](#) no Guia de gerenciamento do Amazon Redshift.

- Java 8

Com exceção do Presto, o Open JDK 1.7 é o padrão JDK usado para todos os aplicativos. No entanto, o Open JDK 1.7 e o 1.8 estão instalados. Para obter mais informações sobre como configurar `JAVA_HOME` para aplicativos, consulte [Configuração de aplicativos para usar Java 8](#).

Problemas conhecidos das versões anteriores que foram resolvidos

- Corrigido um problema de kernel que afetou significativamente o desempenho em volumes otimizados de taxa de transferência HDD (st1) da Amazon EMR no EBS emr-4.6.0.
- Corrigido um problema em que um cluster falharia se alguma zona de HDFS criptografia fosse especificada sem escolher o Hadoop como aplicativo.
- Alterou a política de HDFS gravação padrão de RoundRobin para `AvailableSpaceVolumeChoosingPolicy`. Alguns volumes não foram utilizados adequadamente com a RoundRobin configuração, o que resultou em falhas nos nós principais e na falta HDFS de confiabilidade.

- Corrigido um problema com o EMRFSCLI, que causaria uma exceção ao criar a tabela de metadados padrão do DynamoDB para visualizações consistentes.
- Corrigido um problema de impasse EMRFS que potencialmente ocorria durante operações de renomeação e cópia de várias partes.
- Foi corrigido um problema EMRFS que fazia com que o CopyPart tamanho padrão fosse 5 MB. O padrão agora está definido corretamente como 128 MB.
- Corrigido um problema com a configuração de inicialização do Zeppelin que potencialmente impedia a interrupção do serviço.
- Corrigido um problema com o Spark e o Zeppelin, que impedia você de usar o `s3a://` URI esquema porque ele não `/usr/lib/hadoop/hadoop-aws.jar` estava carregado corretamente em seus respectivos classpath.
- Portado para trás [HUE-2484](#).
- Reportou um [commit](#) do Hue 3.9.0 (não JIRA existe) para corrigir um problema com a amostra do navegador. HBase
- Portado para trás [HIVE-9073](#).

Versões de componentes 4.7.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
emr-ddb	3.1.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.0.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.2.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.4.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.7.1	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.

Componente	Version (Versão)	Descrição
hadoop-client	2.7.2-amzn-2	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.2-amzn-2	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.2-amzn-2	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.2-amzn-2	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-https-server	2.7.2-amzn-2	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.2-amzn-2	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.2-amzn-2	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.2-amzn-2	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.2-amzn-2	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hadoop-yarn-timeline-server	2.7.2-amzn-2	Serviço para recuperar informações atuais e históricas para YARN aplicativos.

Componente	Version (Versão)	Descrição
hbase-hmaster	1.2.1	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.2.1	Serviço para atender uma ou mais HBase regiões.
hbase-client	1.2.1	HBase cliente de linha de comando.
hbase-rest-server	1.2.1	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.2.1	Serviço que fornece um endpoint Thrift para HBase.
hcatalog-client	1.0.0-amzn-5	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	1.0.0-amzn-5	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	1.0.0-amzn-5	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	1.0.0-amzn-5	O cliente da linha de comando do Hive.

Componente	Version (Versão)	Descrição
hive-metastore-server	1.0.0-amzn-5	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	1.0.0-amzn-5	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.7.1-amzn-7	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.12.0	A biblioteca de Machine Learning.
mysql-server	5.5.46	Meu servidor SQL de banco de dados.
oozie-client	4.2.0	O cliente da linha de comando do Oozie.
oozie-server	4.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
phoenix-library	4.7.0- -1,2 HBase	As bibliotecas do phoenix para servidor e cliente
phoenix-query-server	4.7.0- -1,2 HBase	Um servidor leve que fornece JDBC acesso, buffers de protocolo e acesso ao JSON formato do Avatica API

Componente	Version (Versão)	Descrição
presto-coordinator	0.147	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.147	O serviço que executa partes de uma consulta.
pig-client	0.14.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	1.6.1	Os clientes da linha de comando do Spark.
spark-history-server	1.6.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	1.6.1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	1.6.1	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
tez-on-yarn	0.8.3	O YARN aplicativo e as bibliotecas tez.
webserver	2.4.18	HTTP Servidor Apache.

Componente	Version (Versão)	Descrição
zeppelin-server	0.5.6-incubating	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.8	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.8	ZooKeeper cliente de linha de comando.

4.7.0 classificações de configuração

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-4.7.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Hadoop.

Classificações	Descrição
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j	Altere os valores nas propriedades ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j	Altere os valores no arquivo hive-exec-log4j.properties do Hive.

Classificações	Descrição
hive-log4j	Alterar os valores no arquivo hive-log4j.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.

Classificações	Descrição
phoenix-hbase-metrics	Alterar os valores no arquivo hadoop-metrics2-hbase.properties do Phoenix.
phoenix-hbase-site	Alterar os valores no arquivo hbase-site.xml do Phoenix.
phoenix-log4j	Alterar os valores no arquivo log4j.properties do Phoenix.
phoenix-metrics	Alterar os valores no arquivo hadoop-metrics2-phoenix.properties do Phoenix.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.

Classificações	Descrição
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
tez-site	Alterar os valores no arquivo tez-site.xml do Tez.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 4.6.1

4.6.1 versões do aplicativo

As seguintes aplicações são compatíveis com esta versão: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Zeppelin-Sandbox](#) e [ZooKeeper-Sandbox](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-4.6.1	emr-4.6.0	emr-4.5.0	emr-4.4.0
AWS SDK para Java	1.10.27	1.10.27	1.10.27	1.10.27
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
Scala	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
AmazonCloudWatchAgent	-	-	-	-
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

	emr-4.6.1	emr-4.6.0	emr-4.5.0	emr-4.4.0
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	-	-	-	-
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	-	-	-	-

	emr-4.6.1	emr-4.6.0	emr-4.5.0	emr-4.4.0
Zeppelin-Sandbox	0.5.6	0.5.6	0.5.6	0.5.6
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	3.4.8	-	-

Notas de lançamento do 4.6.1

4.6.1 versões de componentes

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
emr-ddb	3.0.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.0.0	Bibliotecas convenientes para o ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
emr-kinesis	3.1.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.3.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.6.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.2-amzn-1	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.2-amzn-1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.2-amzn-1	HDFScliente e biblioteca de linha de comando

Componente	Version (Versão)	Descrição
hadoop-hdfs-namenode	2.7.2-amzn-1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httfs-server	2.7.2-amzn-1	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.2-amzn-1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.2-amzn-1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.2-amzn-1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.2-amzn-1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hbase-hmaster	1.2.0	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.2.0	Serviço para atender uma ou mais HBase regiões.
hbase-client	1.2.0	HBasecliente de linha de comando.

Componente	Version (Versão)	Descrição
hbase-rest-server	1.2.0	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.2.0	Serviço que fornece um endpoint Thrift para HBase
hcatalog-client	1.0.0-amzn-4	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	1.0.0-amzn-4	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	1.0.0-amzn-4	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	1.0.0-amzn-4	O cliente da linha de comando do Hive.
hive-metastore-server	1.0.0-amzn-4	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	1.0.0-amzn-4	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.7.1-amzn-6	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop

Componente	Version (Versão)	Descrição
mahout-client	0.11.1	A biblioteca de Machine Learning.
mysql-server	5.5	Meu servidor SQL de banco de dados.
oozie-client	4.2.0	O cliente da linha de comando do Oozie.
oozie-server	4.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
presto-coordinator	0.143	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.143	O serviço que executa partes de uma consulta.
pig-client	0.14.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	1.6.1	Os clientes da linha de comando do Spark.
spark-history-server	1.6.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	1.6.1	Mecanismo de execução na memória para YARN.

Componente	Version (Versão)	Descrição
spark-yarn-slave	1.6.1	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
webserver	2.4	HTTPServidor Apache.
zeppelin-server	0.5.6-incubating	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.8	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.8	ZooKeeper cliente de linha de comando.

4.6.1 classificações de configuração

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

classificações do emr-4.6.1

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.

Classificações	Descrição
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase hbase-log4j.properties.
hbase-metrics	Altere os valores no arquivo HBase hadoop-metrics2-hbase.properties do.
hbase-policy	Altere os valores no HBase arquivo hbase-policy.xml do.
hbase-site	Altere os valores no HBase arquivo hbase-site.xml do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site.xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.

Classificações	Descrição
hcatalog-webhcat-log4j	Altere os valores nas propriedades ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j	Altere os valores no arquivo hive-exec-log4j.properties do Hive.
hive-log4j	Alterar os valores no arquivo hive-log4j.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.

Classificações	Descrição
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.

Classificações	Descrição
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 4.6.0

Versões do aplicativo 4.6.0

As seguintes aplicações são compatíveis com esta versão: [Ganglia](#), [HBase](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#), [Zeppelin-Sandbox](#) e [ZooKeeper-Sandbox](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-4.6.0	emr-4.5.0	emr-4.4.0	emr-4.3.0
AWS SDK para Java	1.10.27	1.10.27	1.10.27	1.10.27
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
Scala	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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	-	-	-	-

	emr-4.6.0	emr-4.5.0	emr-4.4.0	emr-4.3.0
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	-	-	-	-
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

	emr-4.6.0	emr-4.5.0	emr-4.4.0	emr-4.3.0
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	3.4.8	-	-	-

Notas de lançamento da 4.6.0

As notas de lançamento a seguir incluem informações para a versão EMR 4.6.0 da Amazon.

- Adicionado HBase 1.2.0
- Adicionado o ZooKeeper-Sandbox 3.4.8
- Atualizado para Presto-Sandbox 0.143
- Os EMR lançamentos da Amazon agora são baseados no Amazon Linux 2016.03.0. Para ter mais informações, consulte <https://aws.amazon.com/amazon-linux-ami/2016.03-release-notes/>.
- Problema que afeta os tipos de volume com taxa de transferência otimizada HDD (st1) EBS

Um problema nas versões 4.2 e superiores do kernel Linux afeta significativamente o desempenho nos volumes Throughput Optimized HDD (st1) EBS do. EMR Esta versão (emr-4.6.0) usa uma versão do kernel 4.4.5 e, portanto, é afetada. Portanto, recomendamos não usar o emr-4.6.0 se você quiser usar volumes st1. EBS Você pode usar o emr-4.5.0 ou EMR versões anteriores da Amazon com st1 sem impacto. Além disso, fornecemos a correção com futuras versões.

- Padrões Python

O Python 3.4 agora está instalado por padrão, mas o Python 2.7 permanece como o sistema padrão. Você pode configurar o Python 3.4 como o padrão do sistema usando uma ação de bootstrap; você pode usar a configuração API para definir `PYSPARK_PYTHON` export to `/usr/bin/python3.4` na `spark-env` classificação para afetar a versão do Python usada por. PySpark

- Java 8

Com exceção do Presto, o Open JDK 1.7 é o padrão JDK usado para todos os aplicativos. No entanto, o Open JDK 1.7 e o 1.8 estão instalados. Para obter mais informações sobre como configurar `JAVA_HOME` para aplicativos, consulte [Configuração de aplicativos para usar Java 8](#).

Problemas conhecidos das versões anteriores que foram resolvidos

- Corrigido um problema em que, às vezes, ocorria uma falha aleatória no provisionamento de aplicativos devido a uma senha gerada.
- Anteriormente, `mysqld` estava instalado em todos os nós. Agora, ele só está instalado na instância principal e somente se o aplicativo escolhido incluir `mysql-server` como componente. Atualmente, os seguintes aplicativos incluem o `mysql-server` componente: HiveHCatalog, Hue, Presto-Sandbox e Sqoop-Sandbox.
- Alterado do padrão `32 yarn.scheduler.maximum-allocation-vcores` para 80, o que corrige um problema introduzido no `emr-4.4.0` que ocorre principalmente com o Spark ao usar a `maximizeResourceAllocation` opção em um cluster cujo tipo de instância principal é um dos poucos tipos de instância grandes que têm YARN vcores definidos acima de 32; ou seja, `c4.8xlarge`, `cc2.8xlarge`, `hs1.8xlarge`, `i2.8xlarge`, `m2.4xlarge`, `r3.8xlarge`, `d2.8xlarge` grande ou `m4.10xlarge` foram afetados por esse problema.
- O `s3-dist-cp` agora é usado para todas as indicações ao EMRFS Amazon S3 e não é mais usado em um diretório temporário. HDFS
- Corrigido um problema com o tratamento de exceções para os multipart uploads de criptografia no lado do cliente.
- Adicionada uma opção para permitir que os usuários alterem a classe de armazenamento do Amazon S3. Por padrão, essa configuração é `STANDARD`. A configuração da classificação de configuração `emrfs-site` é `fs.s3.storageClass` e os valores possíveis são `STANDARD`, `STANDARD_IA` e `REDUCED_REDUNDANCY`. Para obter mais informações sobre classes de armazenamento, consulte [Classes de armazenamento](#) no Guia do usuário do Amazon Simple Storage Service.

Versões de componentes 4.6.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`.

O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado myapp-component com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como 2.2-amzn-2.

Componente	Version (Versão)	Descrição
emr-ddb	3.0.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.0.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.1.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.3.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.6.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas

Componente	Version (Versão)	Descrição
		coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.2-amzn-1	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.2-amzn-1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.2-amzn-1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.2-amzn-1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-https-server	2.7.2-amzn-1	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.2-amzn-1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.2-amzn-1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.2-amzn-1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.2-amzn-1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.

Componente	Version (Versão)	Descrição
hbase-hmaster	1.2.0	Serviço para um HBase cluster responsável pela coordenação das regiões e execução de comandos administrativos.
hbase-region-server	1.2.0	Serviço para atender uma ou mais HBase regiões.
hbase-client	1.2.0	HBase cliente de linha de comando.
hbase-rest-server	1.2.0	Serviço que fornece um RESTful HTTP endpoint para HBase.
hbase-thrift-server	1.2.0	Serviço que fornece um endpoint Thrift para HBase.
hcatalog-client	1.0.0-amzn-4	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	1.0.0-amzn-4	Fornecimento de serviços HCatalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	1.0.0-amzn-4	HTTP endpoint fornecendo uma REST interface para HCatalog.
hive-client	1.0.0-amzn-4	O cliente da linha de comando do Hive.

Componente	Version (Versão)	Descrição
hive-metastore-server	1.0.0-amzn-4	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	1.0.0-amzn-4	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.7.1-amzn-6	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.11.1	A biblioteca de Machine Learning.
mysql-server	5.5	Meu servidor SQL de banco de dados.
oozie-client	4.2.0	O cliente da linha de comando do Oozie.
oozie-server	4.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
presto-coordinator	0.143	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.143	O serviço que executa partes de uma consulta.
pig-client	0.14.0-amzn-0	O cliente da linha de comando do Pig.

Componente	Version (Versão)	Descrição
spark-client	1.6.1	Os clientes da linha de comando do Spark.
spark-history-server	1.6.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	1.6.1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	1.6.1	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
webserver	2.4	HTTP Servidor Apache.
zeppelin-server	0.5.6-incubating	O notebook baseado na web que permite a análise de dados interativa.
zookeeper-server	3.4.8	O serviço centralizado de manutenção de informações de configuração, nomenclatura, fornecimento de sincronização distribuída, e fornecimento de serviços de grupo.
zookeeper-client	3.4.8	ZooKeeper cliente de linha de comando.

4.6.0 classificações de configuração

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-4.6.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Hadoop.
hbase-env	Mude os valores no HBase ambiente.
hbase-log4j	Altere os valores no arquivo HBase <code>hbase-log4j.properties</code> .
hbase-metrics	Altere os valores no arquivo HBase <code>hadoop-metrics2-hbase.properties</code> do.
hbase-policy	Altere os valores no HBase arquivo <code>hbase-policy.xml</code> do.
hbase-site	Altere os valores no HBase arquivo <code>hbase-site.xml</code> do.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's <code>hdfs-site.xml</code> .

Classificações	Descrição
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j	Altere os valores nas propriedades ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j	Altere os valores no arquivo hive-exec-log4j.properties do Hive.
hive-log4j	Alterar os valores no arquivo hive-log4j.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.

Classificações	Descrição
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.

Classificações	Descrição
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.
zookeeper-config	Altere os valores no ZooKeeper arquivo zoo.cfg do.
zookeeper-log4j	Altere os valores no ZooKeeper arquivo log4j.properties.

Amazon EMR versão 4.5.0

Versões do aplicativo 4.5.0

As seguintes aplicações são compatíveis com esta versão: [Ganglia](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#) e [Zeppelin-Sandbox](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-4.5.0	emr-4.4.0	emr-4.3.0	emr-4.2.0
AWS SDK para Java	1.10.27	1.10.27	1.10.27	1.10.27
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
Scala	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.7.2	3.6.0
HBase	-	-	-	-
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	-	-	-	-

	emr-4.5.0	emr-4.4.0	emr-4.3.0	emr-4.2.0
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	-	-
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-

	emr-4.5.0	emr-4.4.0	emr-4.3.0	emr-4.2.0
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.6	0.5.6	0.5.5	0.5.5
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-

Notas de lançamento da versão 4.5.0

As notas de lançamento a seguir incluem informações para a versão EMR 4.5.0 da Amazon.

Data do release: 4 de abril de 2016

Recursos

- Atualizado para Spark 1.6.1
- Atualizado para Hadoop 2.7.2
- Atualizado para Presto 0.140
- Foi adicionado AWS KMS suporte para criptografia do lado do servidor Amazon S3.

Problemas conhecidos das versões anteriores que foram resolvidos

- Corrigido um problema em que os servidores My SQL e Apache não iniciavam após a reinicialização de um nó.
- Corrigido um problema em IMPORT que não funcionava corretamente com tabelas não particionadas armazenadas no Amazon S3
- Corrigido um problema em que o Presto exigia que o diretório de preparo fosse /mnt/tmp em vez de /tmp ao gravar em tabelas do Hive.

Versões de componentes 4.5.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion-amzn-EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>emr-ddb</code>	3.0.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.0.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.1.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.2.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
<code>emrfs</code>	2.5.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
<code>ganglia-monitor</code>	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop,

Componente	Version (Versão)	Descrição
		juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.2-amzn-0	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.2-amzn-0	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.2-amzn-0	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.2-amzn-0	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-https-server	2.7.2-amzn-0	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.7.2-amzn-0	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API

Componente	Version (Versão)	Descrição
hadoop-mapred	2.7.2-amzn-0	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.2-amzn-0	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.2-amzn-0	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hcatalog-client	1.0.0-amzn-4	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	1.0.0-amzn-4	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	1.0.0-amzn-4	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	1.0.0-amzn-4	O cliente da linha de comando do Hive.
hive-metastore-server	1.0.0-amzn-4	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL

Componente	Version (Versão)	Descrição
hive-server	1.0.0-amzn-4	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.7.1-amzn-5	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.11.1	A biblioteca de Machine Learning.
mysql-server	5.5	Meu servidor SQL de banco de dados.
oozie-client	4.2.0	O cliente da linha de comando do Oozie.
oozie-server	4.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
presto-coordinator	0.140	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.140	O serviço que executa partes de uma consulta.
pig-client	0.14.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	1.6.1	Os clientes da linha de comando do Spark.

Componente	Version (Versão)	Descrição
spark-history-server	1.6.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	1.6.1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	1.6.1	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
webserver	2.4	HTTP Servidor Apache.
zeppelin-server	0.5.6-incubating	O notebook baseado na web que permite a análise de dados interativa.

4.5.0 classificações de configuração

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-4.5.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.

Classificações	Descrição
emrfs-site	Altere EMRFS as configurações.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j	Altere os valores nas propriedades ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j	Altere os valores no arquivo hive-exec-log4j.properties do Hive.
hive-log4j	Alterar os valores no arquivo hive-log4j.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.

Classificações	Descrição
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.

Classificações	Descrição
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.

Amazon EMR versão 4.4.0

Versões do aplicativo 4.4.0

As seguintes aplicações são compatíveis com esta versão: [Ganglia](#), [HCatalog](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Pig](#), [Presto-Sandbox](#), [Spark](#), [Sqoop-Sandbox](#) e [Zeppelin-Sandbox](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-4.4.0	emr-4.3.0	emr-4.2.0	emr-4.1.0
AWS SDK para Java	1.10.27	1.10.27	1.10.27	Não rastreadas
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
Scala	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
AmazonCloudWatchAgent	-	-	-	-
Delta	-	-	-	-
Flink	-	-	-	-
Ganglia	3.7.2	3.7.2	3.6.0	-
HBase	-	-	-	-

	emr-4.4.0	emr-4.3.0	emr-4.2.0	emr-4.1.0
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	-	-	-	-
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	-	-	-

	emr-4.4.0	emr-4.3.0	emr-4.2.0	emr-4.1.0
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.6	0.5.5	0.5.5	0.6.0-SNAPSHOT
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-

Notas de lançamento da versão 4.4.0

As notas de lançamento a seguir incluem informações para a versão EMR 4.4.0 da Amazon.

Data do release: 14 de março de 2016

Recursos

- Adicionado HCatalog 1.0.0
- Adicionado o Sqoop-Sandbox 1.4.6
- Atualizado para Presto 0.136
- Atualizado para Zeppelin 0.5.6
- Atualizado para Mahout 0.11.1
- Habilitada a `dynamicResourceAllocation` por padrão.
- Adicionada uma tabela de todas as classificações de configuração para a versão. Para obter mais informações, consulte a tabela de classificações de configuração em [Configurar aplicativos](#).

Problemas conhecidos das versões anteriores que foram resolvidos

- Corrigido um problema em que a `maximizeResourceAllocation` configuração não reservava memória suficiente para YARN ApplicationMaster daemons.
- Corrigido um problema encontrado com um personalizadoDNS. Se alguma entrada em `resolve.conf` preceder as entradas personalizadas fornecidas, as entradas personalizadas não serão resolvíveis. Esse comportamento foi afetado por clusters em um em VPC que o servidor de VPC nomes padrão é inserido como a entrada superior em `resolve.conf`.
- Corrigido um problema em que o Python padrão mudou para a versão 2.7 e boto não estava instalado para essa versão.
- Corrigido um problema em que YARN contêineres e aplicativos Spark geravam um arquivo de banco de dados Round Robin (rrd) exclusivo do Ganglia, o que resultava no preenchimento do primeiro disco conectado à instância. Por causa dessa correção, as métricas em nível de YARN contêiner foram desativadas e as métricas em nível de aplicativo do Spark foram desativadas.
- Corrigido um problema no log pusher em que ele excluía todas as pastas de log vazias. O efeito foi que o Hive não CLI conseguiu se registrar porque o carregador de registros estava removendo a user pasta vazia abaixo. `/var/log/hive`
- Corrigido um problema com as importações do Hive, que afetava o particionamento e resultava em um erro durante a importação.
- Corrigido um problema em que EMRFS o `s3-dist-cp` não manipulava adequadamente os nomes de buckets que continham pontos.
- Foi alterado um comportamento EMRFS para que, em buckets com versionamento ativado, o arquivo `_$folder$` marcador não seja criado continuamente, o que pode contribuir para melhorar o desempenho de buckets com versionamento ativado.
- Alterou o comportamento de EMRFS forma que não use arquivos de instruções, exceto nos casos em que a criptografia do lado do cliente está habilitada. Se desejar excluir os arquivos de instrução enquanto usa a criptografia no lado do cliente, você poderá definir a propriedade `emrfs-site.xml fs.s3.cse.cryptoStorageMode.deleteInstructionFiles.enabled`, como verdadeiro.
- A agregação de YARN registros foi alterada para reter os registros no destino da agregação por dois dias. O destino padrão é o HDFS armazenamento em cluster. Se desejar mudar essa duração, altere o valor de `yarn.log-aggregation.retain-seconds` usando a classificação de configuração `yarn-site` quando criar seu cluster. Como sempre, você pode salvar os logs de aplicações no Amazon S3 usando o parâmetro `log-uri` quando criar o cluster.

Patches aplicados

- [HIVE-965](#)
- [HIVE-9183](#)
- [HADOOP-12810](#)

Versões de componentes 4.4.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion*-amzn-*EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>emr-ddb</code>	3.0.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.0.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.1.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.2.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.

Componente	Version (Versão)	Descrição
emrfs	2.4.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.1-amzn-1	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.1-amzn-1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.1-amzn-1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.1-amzn-1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httpfs-server	2.7.1-amzn-1	HTTPendpoint para HDFS operações.

Componente	Version (Versão)	Descrição
hadoop-kms-server	2.7.1-amzn-1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.1-amzn-1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.1-amzn-1	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.1-amzn-1	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hcatalog-client	1.0.0-amzn-3	O cliente da linha de comando 'hcat' para manipular o hcatalog-server.
hcatalog-server	1.0.0-amzn-3	Fornecimento de serviçosH Catalog, uma camada de gerenciamento de tabelas e armazenamento para aplicativos distribuídos.
hcatalog-webhcat-server	1.0.0-amzn-3	HTTPendpoint fornecendo uma REST interface paraHCatalog.
hive-client	1.0.0-amzn-3	O cliente da linha de comando do Hive.

Componente	Version (Versão)	Descrição
hive-metastore-server	1.0.0-amzn-3	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	1.0.0-amzn-3	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.7.1-amzn-5	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.11.1	A biblioteca de Machine Learning.
mysql-server	5.5	Meu servidor SQL de banco de dados.
oozie-client	4.2.0	O cliente da linha de comando do Oozie.
oozie-server	4.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
presto-coordinator	0.136	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.136	O serviço que executa partes de uma consulta.
pig-client	0.14.0-amzn-0	O cliente da linha de comando do Pig.

Componente	Version (Versão)	Descrição
spark-client	1.6.0	Os clientes da linha de comando do Spark.
spark-history-server	1.6.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	1.6.0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	1.6.0	Bibliotecas do Apache Spark necessárias para escravos. YARN
sqoop-client	1.4.6	O cliente da linha de comando do Apache Sqoop.
webserver	2.4	HTTP Servidor Apache.
zeppelin-server	0.5.6-incubating	O notebook baseado na web que permite a análise de dados interativa.

4.4.0 classificações de configuração

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-4.4.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.

Classificações	Descrição
core-site	Alterar os valores no arquivo core-site.xml do Hadoop.
emrfs-site	Altere EMRFS as configurações.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo log4j.properties do Hadoop.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.
hcatalog-env	Mude os valores no HCatalog ambiente.
hcatalog-server-jndi	Altere os valores em HCatalog jndi.properties.
hcatalog-server-proto-hive-site	Altere os valores em HCatalog proto-hive-site .xml.
hcatalog-webhcat-env	Altere os valores no ambiente ebHCat de HCatalog W.
hcatalog-webhcat-log4j	Altere os valores nas propriedades ebHCat log4j.properties de HCatalog W.
hcatalog-webhcat-site	Altere os valores no arquivo webhcat-site.xml de HCatalog W. ebHCat
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j	Altere os valores no arquivo hive-exec-log4j.properties do Hive.
hive-log4j	Alterar os valores no arquivo hive-log4j.properties do Hive.

Classificações	Descrição
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.

Classificações	Descrição
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
sqoop-env	Alterar os valores no ambiente do Sqoop.
sqoop-oraoop-site	Altere os valores no arquivo oraoop-site.xml OraOop do Sqoop.
sqoop-site	Alterar os valores no arquivo sqoop-site.xml do Sqoop.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.

Classificações	Descrição
zeppelin-env	Alterar os valores no ambiente do Zeppelin.

Amazon EMR versão 4.3.0

Versões do aplicativo 4.3.0

As seguintes aplicações são compatíveis com esta versão: [Ganglia](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Pig](#), [Presto-Sandbox](#), [Spark](#) e [Zeppelin-Sandbox](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
AWS SDK para Java	1.10.27	1.10.27	Não rastreadas	Não rastreadas
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
Scala	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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	-	-	-	-

Notas de lançamento do 4.3.0

As notas de lançamento a seguir incluem informações para a versão EMR 4.3.0 da Amazon.

Data do release: 19 de janeiro de 2016

Recursos

- Atualizado para Hadoop 2.7.1
- Atualizado para Spark 1.6.0
- Ganglia atualizado para 3.7.2
- Presto atualizado para 0.130

- A Amazon EMR fez algumas alterações em `spark.dynamicAllocation.enabled` quando está definido como verdadeiro; por padrão, é falso. Quando definida como verdadeiro, isso afeta os padrões definidos pela configuração `maximizeResourceAllocation`:
 - Se `spark.dynamicAllocation.enabled` estiver definida como `true`, `spark.executor.instances` não será definida por `maximizeResourceAllocation`.
 - A configuração `spark.driver.memory` agora é definida com base nos tipos de instância no cluster de maneira semelhante a como `spark.executors.memory` é definida. No entanto, como o aplicativo do driver Spark pode ser executado na instância principal ou em uma das instâncias principais (por exemplo, nos modos YARN cliente e cluster, respectivamente), a `spark.driver.memory` configuração é definida com base no tipo de instância do tipo de instância menor entre esses dois grupos de instâncias.
- A `spark.default.parallelism` configuração agora está definida como o dobro do número de CPU núcleos disponíveis para YARN contêineres. Em versões anteriores, era a metade desse valor.
- Os cálculos da sobrecarga de memória reservada para YARN os processos do Spark foram ajustados para serem mais precisos, resultando em um pequeno aumento na quantidade total de memória disponível para o Spark (ou seja,). `spark.executor.memory`

Problemas conhecidos das versões anteriores que foram resolvidos

- YARNa agregação de registros agora está ativada por padrão.
- Corrigido um problema em que os registros não eram enviados para o bucket de registros do Amazon S3 para o cluster quando a agregação de registros YARN estava ativada.
- YARNos tamanhos de contêineres agora têm um novo mínimo de 32 em todos os tipos de nós.
- Corrigido um problema com o Ganglia que causava E/S de disco excessivas no nó primário em clusters grandes.
- Corrigido um problema que impedia que os logs de aplicações fossem enviados para o Amazon S3 quando um cluster estivesse sendo encerrado.
- Corrigido um problema EMRFS CLI que fazia com que certos comandos falhassem.
- Corrigido um problema com o Zeppelin que impedia que dependências fossem carregadas no subjacente. `SparkContext`
- Corrigido um problema resultante da emissão de um redimensionamento para tentar adicionar instâncias.

- Corrigido um problema no Hive em que o CREATE TABLE AS SELECT fazia chamadas de lista excessivas para o Amazon S3.
- Corrigido um problema em que clusters grandes não provisionavam corretamente quando o Hue, o Oozie e o Ganglia estivessem instalados.
- Corrigido um problema no s3-dist-cp em que retornava um código de saída zero, mesmo se houvesse falha com um erro.

Versões de componentes 4.3.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>emr-ddb</code>	3.0.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.0.0	Bibliotecas convenientes para o ecossistema do Hadoop.
<code>emr-kinesis</code>	3.1.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
<code>emr-s3-dist-cp</code>	2.1.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.

Componente	Version (Versão)	Descrição
emrfs	2.3.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
ganglia-monitor	3.7.2	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.7.2	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.7.1	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.7.1-amzn-0	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.7.1-amzn-0	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.7.1-amzn-0	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.7.1-amzn-0	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httpfs-server	2.7.1-amzn-0	HTTPendpoint para HDFS operações.

Componente	Version (Versão)	Descrição
hadoop-kms-server	2.7.1-amzn-0	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.7.1-amzn-0	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.7.1-amzn-0	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.7.1-amzn-0	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hive-client	1.0.0-amzn-2	O cliente da linha de comando do Hive.
hive-metastore-server	1.0.0-amzn-2	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	1.0.0-amzn-2	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.7.1-amzn-5	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.11.0	A biblioteca de Machine Learning.

Componente	Version (Versão)	Descrição
mysql-server	5.5	Meu servidor SQL de banco de dados.
oozie-client	4.2.0	O cliente da linha de comando do Oozie.
oozie-server	4.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
presto-coordinator	0.130	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.130	O serviço que executa partes de uma consulta.
pig-client	0.14.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	1.6.0	Os clientes da linha de comando do Spark.
spark-history-server	1.6.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	1.6.0	Mecanismo de execução na memória para YARN.
spark-yarn-slave	1.6.0	Bibliotecas do Apache Spark necessárias para escravos. YARN

Componente	Version (Versão)	Descrição
webserver	2.4	HTTPServidor Apache.
zeppelin-server	0.5.5-incubating-amzn-1	O notebook baseado na web que permite a análise de dados interativa.

4.3.0 classificações de configuração

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-4.3.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Hadoop.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's <code>hdfs-site.xml</code> .
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j	Altere os valores no arquivo <code>hive-exec-log4j.properties</code> do Hive.

Classificações	Descrição
hive-log4j	Alterar os valores no arquivo hive-log4j.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.

Classificações	Descrição
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.

Amazon EMR versão 4.2.0

Versões do aplicativo 4.2.0

As seguintes aplicações são compatíveis com esta versão: [Ganglia](#), [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Pig](#), [Presto-Sandbox](#), [Spark](#) e [Zeppelin-Sandbox](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
AWS SDK para Java	1.10.27	1.10.27	Não rastreadas	Não rastreadas
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
Scala	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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	-	-	-	-

Notas de lançamento da versão 4.2.0

As notas de lançamento a seguir incluem informações para a versão EMR 4.2.0 da Amazon.

Data do release: 18 de novembro de 2015

Recursos

- Adicionado o suporte ao Ganglia
- Atualizado para Spark 1.5.2
- Atualizado para Presto 0.125
- Oozie atualizado para 4.2.0
- Zeppelin atualizado para 0.5.5
- Atualizou o AWS SDK for Java para 1.10.27

Problemas conhecidos das versões anteriores que foram resolvidos

- Corrigido um problema EMRFS CLI em que ele não usava o nome padrão da tabela de metadados.

- Corrigido um problema encontrado ao usar tabelas ORC apoiadas por -back no Amazon S3.
- Corrigido um problema encontrado com uma divergência de versão do Python na configuração do Spark.
- Corrigido um problema quando o status de um YARN nó não era reportado devido a DNS problemas com clusters em umVPC.
- Corrigido um problema encontrado ao YARN descomissionar nós, resultando em aplicativos suspensos ou na incapacidade de programar novos aplicativos.
- Corrigido um problema encontrado quando os clusters terminavam com o status TIMED _ OUT _STARTING.
- Corrigido um problema encontrado ao incluir a dependência do EMRFS Scala em outras compilações. A dependência Scala foi removida.

Versões do componente 4.2.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário *CommunityVersion-amzn-EmrVersion*. O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
<code>emr-ddb</code>	3.0.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
<code>emr-goodies</code>	2.0.0	Bibliotecas convenientes para o ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
emr-kinesis	3.1.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.0.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.2.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
ganglia-monitor	3.6.0	O agente incorporado do Ganglia para aplicativos do ecossistema do Hadoop, juntamente com o agente de monitoramento do Ganglia.
ganglia-metadata-collector	3.6.0	O coletor de metadados do Ganglia para agregar métricas de agentes de monitoramento do Ganglia.
ganglia-web	3.5.10	O aplicativo web para visualizar as métricas coletadas pelo coletor de metadados do Ganglia.
hadoop-client	2.6.0-amzn-2	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.6.0-amzn-2	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.6.0-amzn-2	HDFScliente e biblioteca de linha de comando

Componente	Version (Versão)	Descrição
hadoop-hdfs-namenode	2.6.0-amzn-2	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httfs-server	2.6.0-amzn-2	HTTPendpoint para HDFS operações.
hadoop-kms-server	2.6.0-amzn-2	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.6.0-amzn-2	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.6.0-amzn-2	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.6.0-amzn-2	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hive-client	1.0.0-amzn-1	O cliente da linha de comando do Hive.
hive-metastore-server	1.0.0-amzn-1	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	1.0.0-amzn-1	O serviço que aceita as consultas do Hive como solicitações da web.

Componente	Version (Versão)	Descrição
hue-server	3.7.1-amzn-5	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop
mahout-client	0.11.0	A biblioteca de Machine Learning.
mysql-server	5.5	Meu servidor SQL de banco de dados.
oozie-client	4.2.0	O cliente da linha de comando do Oozie.
oozie-server	4.2.0	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
presto-coordinator	0.125	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.125	O serviço que executa partes de uma consulta.
pig-client	0.14.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	1.5.2	Os clientes da linha de comando do Spark.
spark-history-server	1.5.2	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.

Componente	Version (Versão)	Descrição
spark-on-yarn	1.5.2	Mecanismo de execução na memória para YARN.
spark-yarn-slave	1.5.2	Bibliotecas do Apache Spark necessárias para escravos. YARN
webserver	2.4	HTTP Servidor Apache.
zeppelin-server	0.5.5-incubating-amzn-0	O notebook baseado na web que permite a análise de dados interativa.

4.2.0 classificações de configuração

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-4.2.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Hadoop.

Classificações	Descrição
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's hdfs-site.xml.
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j	Altere os valores no arquivo hive-exec-log4j.properties do Hive.
hive-log4j	Alterar os valores no arquivo hive-log4j.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.

Classificações	Descrição
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
presto-connector-hive	Alterar os valores no arquivo hive.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
spark-metrics	Alterar os valores no arquivo metrics.properties do Spark.
yarn-env	Mude os valores no YARN ambiente.

Classificações	Descrição
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.

Amazon EMR versão 4.1.0

Versões do aplicativo 4.1.0

As seguintes aplicações são compatíveis com esta versão: [Hadoop](#), [Hive](#), [Hue](#), [Mahout](#), [Oozie-Sandbox](#), [Pig](#), [Presto-Sandbox](#), [Spark](#) e [Zeppelin-Sandbox](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
AWS SDK para Java	1.10.27	1.10.27	Não rastreadas	Não rastreadas
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
Scala	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
AmazonCloudWatchAgent	-	-	-	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
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	-	-	-	-
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	-	-	-	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
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	-	-	-	-

Notas de lançamento do 4.1.0

Versões do componente 4.1.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`.

O *EmrVersion* começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado myapp-component com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como 2.2-amzn-2.

Componente	Version (Versão)	Descrição
emr-ddb	3.0.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.0.0	Bibliotecas convenientes para o ecossistema do Hadoop.
emr-kinesis	3.1.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.0.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.1.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
hadoop-client	2.6.0-amzn-1	Cientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.6.0-amzn-1	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-library	2.6.0-amzn-1	HDFScliente e biblioteca de linha de comando
hadoop-hdfs-namenode	2.6.0-amzn-1	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.

Componente	Version (Versão)	Descrição
hadoop-https-server	2.6.0-amzn-1	HTTP endpoint para HDFS operações.
hadoop-kms-server	2.6.0-amzn-1	Servidor de gerenciamento de chaves criptográficas baseado no Hadoop. KeyProvider API
hadoop-mapred	2.6.0-amzn-1	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.6.0-amzn-1	YARN serviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resource-manager	2.6.0-amzn-1	YARN serviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.
hive-client	1.0.0-amzn-1	O cliente da linha de comando do Hive.
hive-metastore-server	1.0.0-amzn-1	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	1.0.0-amzn-1	O serviço que aceita as consultas do Hive como solicitações da web.
hue-server	3.7.1-amzn-4	O aplicativo web para analisar dados usando aplicativos do ecossistema do Hadoop

Componente	Version (Versão)	Descrição
mahout-client	0.11.0	A biblioteca de Machine Learning.
mysql-server	5.5	Meu servidor SQL de banco de dados.
oozie-client	4.0.1	O cliente da linha de comando do Oozie.
oozie-server	4.0.1	O serviço que aceita solicitações de fluxo de trabalho do Oozie.
presto-coordinator	0.119	O serviço que aceita consultas e gerencia a execução de consultas entre os presto-workers.
presto-worker	0.119	O serviço que executa partes de uma consulta.
pig-client	0.14.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	1.5.0	Os clientes da linha de comando do Spark.
spark-history-server	1.5.0	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	1.5.0	Mecanismo de execução na memória para YARN.

Componente	Version (Versão)	Descrição
spark-yarn-slave	1.5.0	Bibliotecas do Apache Spark necessárias para escravos. YARN
zeppelin-server	0.6.0-incubando- SNAPSHOT	O notebook baseado na web que permite a análise de dados interativa.

4.1.0 classificações de configuração

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-4.1.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Hadoop.
hdfs-encryption-zones	Configure zonas de HDFS criptografia.
hdfs-site	Altere os valores em HDFS's <code>hdfs-site.xml</code> .
hive-env	Alterar os valores no ambiente do Hive.

Classificações	Descrição
hive-exec-log4j	Altere os valores no arquivo hive-exec-log4j.properties do Hive.
hive-log4j	Alterar os valores no arquivo hive-log4j.properties do Hive.
hive-site	Alterar os valores no arquivo hive-site.xml do Hive.
hue-ini	Alterar os valores no arquivo ini do Hue
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo httpfs-site.xml do Hadoop.
hadoop-kms-acls	Alterar os valores no arquivo kms-acls.xml do Hadoop.
hadoop-kms-env	Altere os valores no ambiente HadoopKMS.
hadoop-kms-log4j	Alterar os valores no arquivo kms-log4j.properties do Hadoop.
hadoop-kms-site	Alterar os valores no arquivo kms-site.xml do Hadoop.
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
oozie-env	Alterar os valores no ambiente do Oozie.
oozie-log4j	Alterar os valores no arquivo oozie-log4j.properties do Oozie.

Classificações	Descrição
oozie-site	Alterar os valores no arquivo oozie-site.xml do Oozie.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
presto-log	Alterar os valores no arquivo log.properties do Presto.
presto-config	Alterar os valores no arquivo config.properties do Presto.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.
zeppelin-env	Alterar os valores no ambiente do Zeppelin.

Amazon EMR versão 4.0.0

Versões do aplicativo 4.0.0

As seguintes aplicações são compatíveis com esta versão: [Hadoop](#), [Hive](#), [Mahout](#), [Pig](#) e [Spark](#).

A tabela abaixo lista as versões do aplicativo disponíveis nesta versão da Amazon EMR e as versões do aplicativo nas três EMR versões anteriores da Amazon (quando aplicável).

Para obter um histórico abrangente das versões do aplicativo para cada lançamento da AmazonEMR, consulte os seguintes tópicos:

- [Versões do aplicativo nas versões EMR 7.x da Amazon](#)
- [Versões de aplicações nas versões 6.x do Amazon EMR](#)
- [Versões de aplicações nas versões 5.x do Amazon EMR](#)
- [Versões de aplicações nas versões 4.x do Amazon EMR](#)

Informações da versão da aplicação

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
AWS SDK para Java	1.10.27	1.10.27	Não rastreadas	Não rastreadas
Python	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
Scala	Não rastreadas	Não rastreadas	Não rastreadas	Não rastreadas
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	-	-	-	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
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	-	-	-	-
TensorFlow	-	-	-	-
Tez	-	-	-	-
Trino (PrestoSQL)	-	-	-	-

	emr-4.3.0	emr-4.2.0	emr-4.1.0	emr-4.0.0
Zeppelin	-	-	-	-
Zeppelin-Sandbox	0.5.5	0.5.5	0.6.0-SNAPSHOT	-
ZooKeeper	-	-	-	-
ZooKeeper-Sandbox	-	-	-	-

Notas de lançamento da 4.0.0

Versões do componente 4.0.0

Os componentes que a Amazon EMR instala com esta versão estão listados abaixo. Alguns são instalados como parte de pacotes de aplicativos de big data. Outros são exclusivos da Amazon EMR e são instalados para processos e recursos do sistema. Eles normalmente começam com `emr` ou `aws`. Os pacotes de aplicativos de big data na versão mais recente da Amazon EMR geralmente são a versão mais recente encontrada na comunidade. Disponibilizamos lançamentos comunitários na Amazon o mais EMR rápido possível.

Alguns componentes na Amazon EMR são diferentes das versões da comunidade. Esses componentes tem um rótulo de versão no formulário `CommunityVersion-amzn-EmrVersion`. O `EmrVersion` começa em 0. Por exemplo, se um componente de comunidade de código aberto nomeado `myapp-component` com a versão 2.2 tiver sido modificado três vezes para inclusão em diferentes EMR versões da Amazon, sua versão de lançamento será listada como `2.2-amzn-2`.

Componente	Version (Versão)	Descrição
emr-ddb	3.0.0	O conector do Amazon DynamoDB para aplicativos do ecossistema do Hadoop.
emr-goodies	2.0.0	Bibliotecas convenientes para o ecossistema do Hadoop.

Componente	Version (Versão)	Descrição
emr-kinesis	3.0.0	O conector do Amazon Kinesis para aplicativos do ecossistema do Hadoop.
emr-s3-dist-cp	2.0.0	Cópia distribuída otimizada de aplicativos para o Amazon S3.
emrfs	2.0.0	O conector do Amazon S3 para aplicações do ecossistema do Hadoop.
hadoop-client	2.6.0-amzn-0	Clientes da linha de comando do Hadoop, como 'hdfs', 'hadoop', ou 'yarn'.
hadoop-hdfs-datanode	2.6.0-amzn-0	HDFSserviço em nível de nó para armazenar blocos.
hadoop-hdfs-namenode	2.6.0-amzn-0	HDFSserviço para rastrear nomes de arquivos e localizações de blocos.
hadoop-httfs-server	2.6.0-amzn-0	HTTPendpoint para HDFS operações.
hadoop-mapred	2.6.0-amzn-0	MapReduce bibliotecas de mecanismos de execução para executar um MapReduce aplicativo.
hadoop-yarn-nodemanager	2.6.0-amzn-0	YARNserviço para gerenciar contêineres em um nó individual.
hadoop-yarn-resourcemanager	2.6.0-amzn-0	YARNserviço para alocar e gerenciar recursos de cluster e aplicativos distribuídos.

Componente	Version (Versão)	Descrição
hive-client	1.0.0-amzn-0	O cliente da linha de comando do Hive.
hive-metastore-server	1.0.0-amzn-0	Serviço para acessar o metastore Hive, um repositório semântico que armazena metadados para operações no Hadoop. SQL
hive-server	1.0.0-amzn-0	O serviço que aceita as consultas do Hive como solicitações da web.
mahout-client	0.10.0	A biblioteca de Machine Learning.
mysql-server	5.5	Meu servidor SQL de banco de dados.
pig-client	0.14.0-amzn-0	O cliente da linha de comando do Pig.
spark-client	1.4.1	Os clientes da linha de comando do Spark.
spark-history-server	1.4.1	A interface de usuário da web para visualizar os eventos registrados por toda a vida útil de um aplicativo Spark concluído.
spark-on-yarn	1.4.1	Mecanismo de execução na memória para YARN.
spark-yarn-slave	1.4.1	Bibliotecas do Apache Spark necessárias para escravos. YARN

Classificações de configuração 4.0.0

As classificações de configuração permitem que você personalize aplicações. Eles geralmente correspondem a um XML arquivo de configuração do aplicativo, como `hive-site.xml`. Para ter mais informações, consulte [Configurar aplicações](#).

Classificações do emr-4.0.0

Classificações	Descrição
capacity-scheduler	Alterar os valores no arquivo <code>capacity-scheduler.xml</code> do Hadoop.
core-site	Alterar os valores no arquivo <code>core-site.xml</code> do Hadoop.
emrfs-site	Altere EMRFS as configurações.
hadoop-env	Alterar os valores no ambiente do Hadoop para todos os componentes do Hadoop.
hadoop-log4j	Alterar os valores no arquivo <code>log4j.properties</code> do Hadoop.
hdfs-site	Altere os valores em HDFS's <code>hdfs-site.xml</code> .
hive-env	Alterar os valores no ambiente do Hive.
hive-exec-log4j	Altere os valores no arquivo <code>hive-exec-log4j.properties</code> do Hive.
hive-log4j	Alterar os valores no arquivo <code>hive-log4j.properties</code> do Hive.
hive-site	Alterar os valores no arquivo <code>hive-site.xml</code> do Hive.
httpfs-env	Mude os valores no HTTPFS ambiente.
httpfs-site	Alterar os valores no arquivo <code>httpfs-site.xml</code> do Hadoop.

Classificações	Descrição
mapred-env	Altere os valores no ambiente do MapReduce aplicativo.
mapred-site	Altere os valores no arquivo mapred-site.xml do MapReduce aplicativo.
pig-properties	Alterar os valores no arquivo pig.properties do Pig.
pig-log4j	Alterar os valores no arquivo log4j.properties do Pig.
spark	Configurações EMR selecionadas pela Amazon para o Apache Spark.
spark-defaults	Alterar os valores no arquivo spark-defaults.conf do Spark.
spark-env	Alterar os valores no ambiente do Spark.
spark-log4j	Alterar os valores no arquivo log4j.properties do Spark.
yarn-env	Mude os valores no YARN ambiente.
yarn-site	Altere os valores no YARN arquivo yarn-site.xml do.

Versões Amazon EMR 2.x e AMI 3.x

Note

AWS está atualizando a TLS configuração de todos os AWS API endpoints para uma versão mínima de TLS 1.2. As EMR versões 3.10 e inferiores da Amazon suportam apenas conexões TLS 1.0/1.1. Depois de 4 de dezembro de 2023, você não poderá criar clusters com o Amazon EMR 3.10 e versões anteriores.

Se você usa o Amazon EMR 3.10 ou inferior, recomendamos que teste e migre imediatamente suas cargas de trabalho para a versão mais recente da Amazon. EMR Para obter mais informações, consulte o [Blog de segurança da AWS](#).

As versões EMR 2.x e 3.x da Amazon, chamadas de AMI versões, são disponibilizadas para soluções preexistentes que as exigem por motivos de compatibilidade. Não recomendamos a criação de novos clusters ou novas soluções com essas versões. Elas não têm recursos de versões mais recentes e incluem pacotes de aplicativos desatualizados.

Recomendamos que você crie soluções usando a EMR versão mais recente da Amazon.

O escopo das diferenças entre as versões das séries 2.x e 3.x e as versões recentes da Amazon EMR é significativo. Essas diferenças vão desde como criar e configurar um cluster para as portas e a estrutura do diretório dos aplicativos no cluster.

Esta seção tenta abordar as diferenças mais significativas da AmazonEMR, bem como as diferenças específicas de configuração e gerenciamento de aplicativos. Ela não é abrangente. Se você criar e usar clusters nas séries 2.x ou 3.x, poderá encontrar diferenças não abordadas nesta seção.

Tópicos

- [Criação de um cluster com AMI versões anteriores da Amazon EMR](#)
- [Instalando aplicativos com AMI versões anteriores da Amazon EMR](#)
- [Personalização da configuração de clusters e aplicativos com AMI versões anteriores da Amazon EMR](#)
- [Especificações do aplicativo Hive para AMI versões anteriores da Amazon EMR](#)
- [HBaseespecificações do aplicativo para AMI versões anteriores da Amazon EMR](#)
- [Especificações do aplicativo Pig para AMI versões anteriores da Amazon EMR](#)
- [Especificações do aplicativo Spark com AMI versões anteriores da Amazon EMR](#)
- [Diferenças de DistCp utilitários do S3 com AMI versões anteriores da Amazon EMR](#)

Criação de um cluster com AMI versões anteriores da Amazon EMR

As versões EMR 2.x e 3.x da Amazon são referenciadas por versão. AMI Com a EMR versão 4.0.0 e posterior da Amazon, os lançamentos são referenciados por versão de lançamento, usando um

rótulo de lançamento como. `emr-5.11.0` Essa alteração é mais aparente quando você cria um cluster usando o AWS CLI ou programaticamente.

Ao usar o AWS CLI para criar um cluster usando uma versão de AMI lançamento, use a `--ami-version` opção, por exemplo, `--ami-version 3.11.0`. Muitas opções, recursos e aplicativos introduzidos no Amazon EMR 4.0.0 e versões posteriores não estão disponíveis quando você especifica um `--ami-version`. Para obter mais informações, consulte [create-cluster](#) na AWS CLI Command Reference.

O AWS CLI comando de exemplo a seguir inicia um cluster usando uma AMI versão.

Note

Os caracteres de continuação de linha do Linux (`\`) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (`^`).

```
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"]
```

Programaticamente, todas as versões de EMR lançamento da Amazon usam a `RunJobFlowRequest` ação no EMR API para criar clusters. O exemplo de código Java a seguir cria um cluster usando a AMI versão 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");
```


A chamada `RunJobFlowRequest` usa um rótulo de versão:

```
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");
```

Configurar o tamanho do cluster

Durante a execução do cluster, o Hadoop determina o número de tarefas de mapeador e reduzir necessárias para processar os dados. Clusters maiores devem ter mais tarefas para fazer melhor utilização dos recursos e reduzir o tempo de processamento. Normalmente, um EMR cluster permanece do mesmo tamanho durante todo o cluster; você define o número de tarefas ao criar o cluster. Quando você redimensiona um cluster em execução, pode variar o processamento durante a execução do cluster. Portanto, em vez de usar um número fixo de tarefas, você pode variar o número de tarefas durante a vida útil do cluster. Há duas opções de configuração para ajudar a definir o número ideal de tarefas:

- `mapred.map.tasksperslot`
- `mapred.reduce.tasksperslot`

Você pode definir ambas as opções no arquivo `mapred-conf.xml`. Quando você envia um trabalho para o cluster, o cliente do trabalho verifica o número total de slots de mapeamento e redução em todo o cluster. O cliente do trabalho, em seguida, usa as seguintes equações para definir o número de tarefas:

- `mapred.map.tasks = mapred.map.tasksperslot * slots de mapeamento no cluster`
- `mapred.reduce.tasks = mapred.reduce.tasksperslot * slots de redução no cluster`

O cliente do trabalho lê o parâmetro `tasksperslot` somente se o número de tarefas não estiver configurado. Você pode substituir o número de tarefas a qualquer momento para todos os clusters por meio de uma ação de bootstrap, ou pode fazê-lo individualmente por trabalho, adicionando uma etapa para alterar a configuração.

A Amazon EMR resiste às falhas do nó da tarefa e continua a execução do cluster mesmo se um nó da tarefa ficar indisponível. A Amazon provisiona EMR automaticamente nós de tarefas adicionais para substituir aqueles que falham.

Você pode ter um número diferente de nós de tarefa para cada etapa do cluster. Você também pode adicionar uma etapa em um cluster em execução para modificar o número de nós de tarefa. Como é certo que todas as etapas são executadas sequencialmente por padrão, você pode especificar o número de nós de tarefa em execução para qualquer etapa.

Instalando aplicativos com AMI versões anteriores da Amazon EMR

Ao usar uma AMI versão, os aplicativos são instalados de várias maneiras, inclusive usando o `NewSupportedProducts` parâmetro para a [RunJobFlow](#) ação, usando ações de bootstrap e usando a ação [Step](#).

Personalização da configuração de clusters e aplicativos com AMI versões anteriores da Amazon EMR

A EMR versão 4.0.0 da Amazon introduziu um método simplificado de configuração de aplicativos usando classificações de configuração. Para obter mais informações, consulte [Configurar aplicações](#). Ao usar uma AMI versão, você configura aplicativos usando ações de bootstrap junto com os argumentos que você passa. Por exemplo, as ações `configure-hadoop` e `configure-daemons` bootstrap definem o Hadoop e propriedades YARN específicas do ambiente, como. `--namenode-heap-size` Em versões mais recentes, elas são configuradas usando as classificações de configuração `hadoop-env` e `yarn-env`. Para ações de bootstrap que realizam configurações comuns, consulte o [emr-bootstrap-actions repositório no Github](#).

As tabelas a seguir mapeiam as ações de bootstrap para as classificações de configuração nas EMR versões mais recentes da Amazon.

Hadoop

Nome do arquivo de aplicativo afetado	AMIação de bootstrap da versão	Classificação de configuração
<code>core-site.xml</code>	<code>configure-hadoop -c</code>	<code>core-site</code>
<code>log4j.properties</code>	<code>configure-hadoop -l</code>	<code>hadoop-log4j</code>

Nome do arquivo de aplicativo afetado	Comando de bootstrap da versão	Classificação de configuração
<code>hdfs-site.xml</code>	<code>configure-hadoop -s</code>	<code>hdfs-site</code>
<code>n/a</code>	<code>n/a</code>	<code>hdfs-encryption-zones</code>
<code>mapred-site.xml</code>	<code>configure-hadoop -m</code>	<code>mapred-site</code>
<code>yarn-site.xml</code>	<code>configure-hadoop -y</code>	<code>yarn-site</code>
<code>httpfs-site.xml</code>	<code>configure-hadoop -t</code>	<code>httpfs-site</code>
<code>capacity-scheduler.xml</code>	<code>configure-hadoop -z</code>	<code>capacity-scheduler</code>
<code>yarn-env.sh</code>	<code>configure-daemons --resourcemanager-opts</code>	<code>yarn-env</code>

Hive

Nome do arquivo de aplicativo afetado	Comando de bootstrap da versão	Classificação de configuração
<code>hive-env.sh</code>	<code>n/a</code>	<code>hive-env</code>
<code>hive-site.xml</code>	<code>hive-script --install -hive-site \${MY_HIVE_SITE_FILE}</code>	<code>hive-site</code>
<code>hive-exec-log4j.properties</code>	<code>n/a</code>	<code>hive-exec-log4j</code>
<code>hive-log4j.properties</code>	<code>n/a</code>	<code>hive-log4j</code>

EMRFS

Nome do arquivo de aplicativo afetado	AMIação de bootstrap da versão	Classificação de configuração
emrfs-site.xml	configure-hadoop -e	emrfs-site
n/a	s3get -s s3://custom-provider.jar -d /usr/share/aws/emr/auxlib/	emrfs-site (com nova configuração fs.s3.cse.encryptionMaterialsProvider.uri)

Para obter uma lista de todas as classificações, consulte [Configurar aplicações](#).

Variáveis de ambientes de aplicações

Ao usar uma AMI versão, um `hadoop-user-env.sh` script é usado junto com a ação `configure-daemons bootstrap` para configurar o ambiente Hadoop. O script inclui as seguintes ações:

```
#!/bin/bash
export HADOOP_USER_CLASSPATH_FIRST=true;
echo "HADOOP_CLASSPATH=/path/to/my.jar" >> /home/hadoop/conf/hadoop-user-env.sh
```

Na EMR versão 4.x da Amazon, você faz o mesmo usando a classificação de `hadoop-env` configuração, conforme mostrado no exemplo a seguir:

```
[
  {
    "Classification": "hadoop-env",
    "Properties": {

    },
    "Configurations": [
      {
        "Classification": "export",
        "Properties": {
          "HADOOP_USER_CLASSPATH_FIRST": "true",
          "HADOOP_CLASSPATH": "/path/to/my.jar"
        }
      }
    ]
  }
]
```

```

    }
  ]

```

Como outro exemplo, usar `configure-daemons` e transmitir `--namenode-heap-size=2048` e `--namenode-opts=-XX:GCTimeRatio=19` é equivalente às seguintes classificações de configuração.

```

[
  {
    "Classification": "hadoop-env",
    "Properties": {
    },
    "Configurations": [
      {
        "Classification": "export",
        "Properties": {
          "HADOOP_DATANODE_HEAPSIZE": "2048",
          "HADOOP_NAMENODE_OPTS": "-XX:GCTimeRatio=19"
        }
      }
    ]
  }
]

```

As outras variáveis do ambiente de aplicativos não são mais definidas em `/home/hadoop/.bashrc`. Em vez disso, elas são definidas nos arquivos `/etc/default` por componente ou aplicativo, como por exemplo `/etc/default/hadoop`. Os scripts de wrapper `/usr/bin/` instalados pelo aplicativo também RPMs podem definir variáveis de ambiente adicionais antes de envolver o script bin real.

Portas de serviço

Ao usar uma AMI versão, alguns serviços usam portas personalizadas.

Alterações nas configurações das portas

Configuração	AMI versão 3.x	Padrão de código aberto
<code>fs.default.name</code>	<code>hdfs://IP: 9000 emrDetermined</code>	padrão (<code>hdfs://<i>emrDeterminedIP</i> 8:80 (20)</code>)

Configuração	AMlversão 3.x	Padrão de código aberto
dfs.datanode.address	0.0.0.0:9200	default (0.0.0.0:50010)
dfs.datanode.http.address	0.0.0.0:9102	default (0.0.0.0:50075)
dfs.datanode.https.address	0.0.0.0:9402	default (0.0.0.0:50475)
dfs.datanode.ipc.address	0.0.0.0:9201	default (0.0.0.0:50020)
dfs.http.address	0.0.0.0:9101	default (0.0.0.0:50070)
dfs.https.address	0.0.0.0:9202	default (0.0.0.0:50470)
dfs.secondary.http.address	0.0.0.0:9104	default (0.0.0.0:50090)
yarn.nodemanager.address	0.0.0.0:9103	default (\${yarn.nodemanager.hostname}:0)
yarn.nodemanager.localizer.address	0.0.0.0:9033	default (\${yarn.nodemanager.hostname}:8040)
yarn.nodemanager.webapp.address	0.0.0.0:9035	default (\${yarn.nodemanager.hostname}:8042)
yarn.resourcemanager.addresses	<i>emrDeterminedIP</i> : 9022	default (\${yarn.resourcemanager.hostname}:8032)
yarn.resourcemanager.admin.address	<i>emrDeterminedIP</i> : 9025	default (\${yarn.resourcemanager.hostname}:8033)
yarn.resourcemanager.resource-tracker.address	<i>emrDeterminedIP</i> : 9023	default (\${yarn.resourcemanager.hostname}:8031)
yarn.resourcemanager.scheduler.address	<i>emrDeterminedIP</i> : 9024	default (\${yarn.resourcemanager.hostname}:8030)
yarn.resourcemanager.webapp.address	0.0.0.0:9026	default (\${yarn.resourcemanager.hostname}:8088)
yarn.web-proxy.address	<i>emrDeterminedIP</i> : 9046	default (no-value)

Configuração	AMI versão 3.x	Padrão de código aberto
yarn.resourcemanager.hostname	0.0.0.0 (default)	<i>emrDeterminedIP</i>

Note

A ferramenta *emrDeterminedIP* é um endereço IP gerado pela AmazonEMR.

Usuários

Ao usar uma AMI versão, o usuário `hadoop` executa todos os processos e possui todos os arquivos. Na EMR versão 4.0.0 e posterior da Amazon, os usuários existem no nível do aplicativo e do componente.

Sequência de instalação, artefatos instalados e locais dos arquivos de log

Ao usar uma AMI versão, os artefatos do aplicativo e seus diretórios de configuração são instalados no `/home/hadoop/application` diretório. Por exemplo, se você tivesse o Hive instalado, o diretório seria `/home/hadoop/hive`. Na EMR versão 4.0.0 e posterior da Amazon, os artefatos do aplicativo são instalados no `/usr/lib/application` diretório. Ao usar uma AMI versão, os arquivos de log são encontrados em vários lugares. A tabela a seguir lista os locais.

Alterações nos locais de log no Amazon S3

Daemon ou aplicação	Local do diretório
instance-state	nó/ <i>instance-id</i> /estado da instância/
hadoop-hdfs-namenode	daemons/ <i>instance-id</i> /hadoop-hadoop-namenode.log
hadoop-hdfs-datanode	daemons/ <i>instance-id</i> /hadoop-hadoop-datanode.log
hadoop () ResourceManager	daemons/ <i>instance-id</i> /yarn-hadoop-resource-manager

Daemon ou aplicação	Local do diretório
hadoop-yarn (Proxy Server)	daemons/ <i>instance-id</i> /yarn-hadoop-proxy server
mapred-historyserver	daemons/ <i>instance-id</i> /
httpfs	daemons/ <i>instance-id</i> /httpfs.log
hive-server	nó/ <i>instance-id</i> /hive-server/hive-server.log
hive-metastore	nó/ <i>instance-id</i> /apps/hive.log
Colmeia CLI	nó/ <i>instance-id</i> /apps/hive.log
YARNregistros de usuários de aplicativos e registros de contêineres	task-attempts/
Mahout	N/D
Pig	N/D
spark-historyserver	N/D
Arquivos de histórico de trabalhos do mapreduce	jobs/

Executor de comandos

Ao usar uma AMI versão, muitos scripts ou programas, por exemplo/home/hadoop/contrib/streaming/hadoop-streaming.jar, não são colocados no ambiente do caminho de login do shell, então você precisa especificar o caminho completo ao usar um arquivo jar, como command-runner.jar ou script-runner.jar, para executar os scripts. Ele command-runner.jar está localizado noAMI, portanto, não há necessidade de saber a totalidadeURI, como foi o caso comscript-runner.jar.

Fator de replicação

O fator de replicação permite que você configure quando iniciar um JVM Hadoop. Você pode iniciar um novo Hadoop JVM para cada tarefa, o que proporciona melhor isolamento de tarefas, ou você

pode compartilhar JVMs entre tarefas, proporcionando menor sobrecarga da estrutura. Se você estiver processando muitos arquivos pequenos, faz sentido reutilizá-los JVM várias vezes para amortizar o custo de inicialização. No entanto, se cada tarefa levar muito tempo ou processar uma grande quantidade de dados, você poderá optar por não reutilizá-las JVM para garantir que toda a memória seja liberada para tarefas subsequentes. Ao usar uma AMI versão, você pode personalizar o fator de replicação usando a ação `configure-hadoop bootstrap` para definir a propriedade `mapred.job.reuse.jvm.num.tasks`.

O exemplo a seguir demonstra a configuração do fator de JVM reutilização para reutilização infinita JVM.

Note

Os caracteres de continuação de linha do Linux (`\`) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (`^`).





```
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"]
```

Especificações do aplicativo Hive para AMI versões anteriores da Amazon EMR

Arquivos de log

Usando EMR AMI as versões 2.x e 3.x da Amazon, os registros do Hive são salvos em `/mnt/var/log/apps/`. Para oferecer suporte a versões simultâneas do Hive, a versão do Hive que você executa determina o nome do arquivo de log, conforme mostrado na tabela a seguir.

Versão do Hive	Nome do arquivo de log
0.13.1	hive.log

Versão do Hive	Nome do arquivo de log
	<p> Note</p> <p>A partir desta versão, a Amazon EMR usa um nome de arquivo sem versão, <code>hive.log</code>. Versões secundárias compartilham o mesmo local de log que a versão principal.</p>
0.11.0	<p><code>hive_0110.log</code></p> <p> Note</p> <p>As versões secundárias do Hive 0.11.0, como o Hive 0.11.0.1, compartilham o mesmo local de arquivo de log que o Hive 0.11.0.</p>
0.8.1	<p><code>hive_081.log</code></p> <p> Note</p> <p>As versões secundárias do Hive 0.8.1, como o Hive 0.8.1.1, compartilham o mesmo local de arquivo de log que o Hive 0.8.1.</p>
0.7.1	<p><code>hive_07_1.log</code></p> <p> Note</p> <p>As versões secundárias do Hive 0.7.1, como o Hive 0.7.1.3 e o Hive 0.7.1.4, compartilham o mesmo local de arquivo de log que o Hive 0.7.1.</p>
0.7	<p><code>hive_07.log</code></p>
0,5	<p><code>hive_05.log</code></p>

Versão do Hive	Nome do arquivo de log
0.4	hive.log

Funcionalidade de entrada dividida

Para implementar a funcionalidade de entrada dividida usando versões do Hive anteriores à 0.13.1 (EMRAMIversões da Amazon anteriores à 3.11.0), use o seguinte:

```
hive> set hive.input.format=org.apache.hadoop.hive.q1.io.HiveCombineSplitsInputFormat;
hive> set mapred.min.split.size=100000000;
```

Essa funcionalidade se tornou obsoleta com o Hive 0.13.1. Para obter a mesma funcionalidade de formato de entrada dividido na Amazon EMR AMI versão 3.11.0, use o seguinte:

```
set hive.hadoop.supports.splittable.combineinputformat=true;
```

Portas de serviço do Thrift

Thrift é uma RPC estrutura que define um formato compacto de serialização binária usado para persistir estruturas de dados para análise posterior. Normalmente, o Hive configura o servidor para operar nas seguintes portas.

Versão do Hive	Número da porta
Hive 0.13.1	10000
Hive 0.11.0	10004
Hive 0.8.1	10003
Hive 0.7.1	10002
Hive 0.7	10001
Hive 0.5	10000

Para obter mais informações sobre serviços do Thrift, consulte <http://wiki.apache.org/thrift/>.

Usar o Hive para recuperar partições

A Amazon EMR inclui uma declaração na linguagem de consulta Hive que recupera as partições de uma tabela a partir dos dados da tabela localizados no Amazon S3. O exemplo a seguir mostra isso.

```
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;
```

Os dados e diretórios da partição devem estar na localização especificada na definição de tabela e devem ser nomeados de acordo com a convenção do Hive: por exemplo, dt=2009-01-01.

Note

Depois do Hive 0.13.1, esse recurso tem suporte nativamente usando `msck repair table` e, portanto, `recover partitions` não tem suporte. Para obter mais informações, consulte <https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL>.

Transmitir uma variável do Hive para um script

Para passar uma variável para uma etapa do Hive usando o AWS CLI, digite o seguinte comando, substitua *myKey* com o nome do seu EC2 key pair e substitua *mybucket* com o nome do seu bucket. Neste exemplo, *SAMPLE* é um valor de variável precedido pela opção `-d`. Essa variável é definida no script Hive como: `${SAMPLE}`.

Note

Os caracteres de continuação de linha do Linux (`\`) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (`^`).

```
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/]
```

Especificar um local de metastore externo

O procedimento a seguir mostra como substituir os valores de configuração padrão para a localização da metastore do Hive e iniciar um cluster usando a localização da metastore reconfigurada.

Para criar um metastore localizado fora do cluster EMR

1. Crie um banco de dados My SQL ou Aurora usando a Amazon. RDS

Para obter informações sobre como criar um RDS banco de dados da Amazon, consulte [Introdução à Amazon RDS](#).

2. Modifique seus grupos de segurança para permitir JDBC conexões entre seu banco de dados e o grupo de segurança ElasticMapReduce-Master.

Para obter informações sobre como modificar seus grupos de segurança para acesso, consulte [Grupos de RDS segurança da Amazon](#) no Guia RDS do usuário da Amazon.

3. Defina os valores JDBC de configuração em `hive-site.xml`:
 - a. Crie um arquivo de configuração `hive-site.xml` contendo as seguintes informações:

```
<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 é o DNS endereço da RDS instância da Amazon que executa o banco de dados. *username* e *password* são as credenciais do seu banco de dados. Para obter mais informações sobre como se conectar às instâncias de banco de dados My SQL e Aurora, consulte [Conectando-se a uma instância de banco de dados executando o mecanismo de SQL banco de dados My](#) e [Conectando-se a um cluster de banco de dados Aurora no Guia do usuário](#) da Amazon. RDS

Os JDBC drivers são instalados pela AmazonEMR.

Note

A propriedade "value" não deve conter espaços ou retornos de carro. Eles devem aparecer todos em uma única linha.

- b. Salve o arquivo `hive-site.xml` em um local no Amazon S3, como `s3://mybucket/hive-site.xml`.
4. Crie um cluster especificando a localização do Amazon S3 do arquivo `hive-site.xml` personalizado.

O exemplo de comando a seguir demonstra um AWS CLI comando que faz isso.

Note

Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

```
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"]
```

Conecte-se ao Hive usando JDBC

Para se conectar ao Hive via, é JDBC necessário baixar o JDBC driver e instalar um SQL cliente. O exemplo a seguir demonstra o uso do SQL Workbench/J para se conectar ao Hive usando JDBC

Para baixar JDBC drivers

1. Faça o download e extraia os drivers apropriados para as versões do Hive que você deseja acessar. A versão do Hive difere dependendo da AMI que você escolher ao criar um EMR cluster da Amazon.
 - [Drivers do Hive 0.13.1: _1.0.4.1004.zip JDBC https://amazon-odbc-jdbc-drivers.s3.amazonaws.com/public/AmazonHiveJDBC](https://amazon-odbc-jdbc-drivers.s3.amazonaws.com/public/AmazonHiveJDBC)
 - Controladores do Hive 0.11.0: JDBC <https://mvnrepository.com/artifact/org.apache.hive/hive-jdbc/0.11.0>
 - Controladores do Hive 0.8.1JDBC: <https://mvnrepository.com/artifact/org.apache.hive/hive-jdbc/0.8.1>
2. Instale o SQL Workbench/J. Para obter mais informações, consulte [Instalando e iniciando o SQL Workbench/J no Manual do usuário do SQL Workbench/J](#).
3. Crie um SSH túnel para o nó principal do cluster. A porta de conexão é diferente dependendo da versão do Hive. Comandos de exemplo são fornecidos nas tabelas abaixo para ssh usuários de Linux e TTY comandos Pu para usuários de Windows

SSHComandos Linux

Versão do Hive	Comando
0.13.1	ssh -o ServerAliveInterval=10 -i <i>path-to-key-file</i> -N -L 10000:localhost:10000 hadoop@ <i>master-public-dns-name</i>
0.11.0	ssh -o ServerAliveInterval=10 -i <i>path-to-key-file</i> -N -L 10004:localhost:10004 hadoop@ <i>master-public-dns-name</i>

Versão do Hive	Comando
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>
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>

Configurações de TTY túnel do Windows Pu

Versão do Hive	Configurações do túnel
0.13.1	Porta de origem: 10000 Destino: <i>master-public-dns-name</i> : 10000
0.11.0	Porta de origem: 10004 Destino: <i>master-public-dns-name</i> : 10004
0.8.1	Porta de origem: 10003 Destino: <i>master-public-dns-name</i> : 10003

4. Adicione o JDBC driver ao SQL Workbench.
 - a. Na caixa de diálogo Select Connection Profile (Selecionar perfil de conexão), escolha Manage Drivers (Gerenciar drivers).
 - b. Escolha o ícone Create a new entry (Criar uma nova entrada) (página em branco).
 - c. No campo Name (Nome), digite **Hive JDBC**.
 - d. Em Biblioteca, clique no ícone Selecionar o (s) JAR arquivo (s).
 - e. Selecione JAR os arquivos conforme mostrado na tabela a seguir.

Versão do driver do Hive	JARarquivos a serem adicionados
0.13.1	<pre>hive_metastore.jar hive_service.jar HiveJDBC3.jar libfb303-0.9.0.jar 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>

Versão do driver do Hive	JARarquivos a serem adicionados
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>
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. Na caixa de diálogo Please select one driver (Selecione um driver), selecione um driver de acordo com a tabela a seguir e clique em OK.

Versão do Hive	Nome da classe do driver
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>
0.7	<code>org.apache.hadoop.hive.jdbc.HiveDriver.jar</code>
0,5	<code>org.apache.hadoop.hive.jdbc.HiveDriver.jar</code>

5. Ao retornar à caixa de diálogo Selecionar Perfil de Conexão, verifique se o campo Driver está definido como Hive JDBC e forneça a cadeia de JDBC conexão no URL campo de acordo com a tabela a seguir.

Versão do Hive	JDBCcadeia de conexão
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>

Se seu cluster usa a AMI versão 3.3.1 ou posterior, na caixa de diálogo Selecionar perfil de conexão, digite o **hadoop** campo Nome de usuário.

HBase especificações do aplicativo para AMI versões anteriores da Amazon EMR

HBase Versões suportadas

HBase versão	AMI versão	AWS CLI parâmetros de configuração	HBase detalhes da versão
0.94.18	3.1.0 e posterior	<pre>--ami-version 3.1 --ami-version 3.2 --ami-version 3.3 --applications Name=HBase</pre>	<ul style="list-style-type: none"> Correções de bugs e melhorias.
0.94.7	3.0-3.0.4	<pre>--ami-version 3.0 --applications Name=HBase</pre>	
0.92	2.2 e posterior	<pre>--ami-version 2.2 or later --applications Name=HBase</pre>	

HBase pré-requisitos do cluster

Um cluster criado usando EMR AMI as versões 2.x e 3.x da Amazon deve atender aos seguintes requisitos para HBase

- O AWS CLI (opcional) — Para interagir HBase usando a linha de comando, baixe e instale a versão mais recente do AWS CLI. Para obter mais informações, consulte [Instalar a AWS Command Line Interface](#) no Guia do usuário da AWS Command Line Interface .
- Pelo menos duas instâncias (opcional) — o nó principal do cluster executa o servidor HBase mestre e o Zookeeper, e os nós de tarefas executam os servidores da HBase região. Para obter o melhor desempenho, HBase os clusters devem ser executados em pelo menos duas EC2 instâncias, mas você pode ser HBase executado em um único nó para fins de avaliação.
- Cluster de longa duração — HBase só é executado em clusters de longa duração. Por padrão, o EMR console CLI e a Amazon criam clusters de longa duração.
- Um conjunto de pares de EC2 chaves da Amazon (recomendado) — Para usar o protocolo de rede Secure Shell (SSH) para se conectar ao nó principal e executar comandos HBase shell, você deve usar um par de EC2 chaves da Amazon ao criar o cluster.
- As versões corretas AMI e do Hadoop — atualmente, os HBase clusters são suportados somente no Hadoop 20.205 ou posterior.
- Ganglia (opcional) — Para monitorar as métricas de HBase desempenho, instale o Ganglia ao criar o cluster.
- Um bucket do Amazon S3 para registros (opcional) — Os registros para HBase estão disponíveis no nó principal. Se você quiser que esses logs sejam copiados para o Amazon S3, especifique um bucket do S3 para receber arquivos de log quando criar o cluster.

Criar um cluster com o HBase

A tabela a seguir lista as opções que estão disponíveis ao usar o console para criar um cluster HBase usando uma versão de EMR AMI lançamento da Amazon.

Campo	Ação
Restore from backup (Restauração de backup)	Especifique se deseja pré-carregar o HBase cluster com dados armazenados no Amazon S3.
Backup location (Localização do backup)	Especifique URI onde o backup a partir do qual a restauração reside no Amazon S3.

Campo	Ação
Backup version (Versão do backup)	Opcionalmente, especifique o nome da versão do backup para uso em Backup Location (Localização do backup). Se você deixar esse campo em branco, a Amazon EMR usará o backup mais recente no Local de Backup para preencher o novo HBase cluster.
Schedule Regular Backups (Agendar backups regulares)	Especifique se deseja agendar backups incrementais automáticos. O primeiro backup é um backup completo para criar uma linha de base para futuros backups incrementais.
Consistent backup (Backup consistente)	Especifica se os backups devem ser consistentes. Um backup consistente é aquele que pausa operações de gravação durante o estágio de backup inicial, a sincronização entre nós. Qualquer operações de gravação assim pausada é colocada em uma fila e retomada quando a sincronização é concluída.
Backup frequency (Frequência de backup)	O número de dias/horas/minutos entre os backups agendados.
Backup location (Localização do backup)	O Amazon S3, URI onde os backups são armazenados. O local de backup de cada HBase cluster deve ser diferente para garantir que os backups diferenciados permaneçam corretos.
Backup start time (Hora de início do backup)	Especifique quando o primeiro backup deve ocorrer. Você pode definir isso comonow, o que faz com que o primeiro backup seja iniciado assim que o cluster estiver em execução ou inserir uma data e hora no ISOformato . Por exemplo, 2012-06-15T 20:00 Z definiria o horário de início para 15 de junho de 2012 às 20:00. UTC

O AWS CLI comando de exemplo a seguir inicia um cluster com HBase e outros aplicativos:

Note

Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

```
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
```

Depois que a conexão entre o Hive e os HBase clusters for estabelecida (conforme mostrado no procedimento anterior), você poderá acessar os dados armazenados no HBase cluster criando uma tabela externa no Hive.

O exemplo a seguir, quando executado a partir do prompt do Hive, cria uma tabela externa que faz referência aos dados armazenados em uma HBase tabela chamada `inputTable`. Em seguida, você pode fazer referência `inputTable` nas instruções do Hive para consultar e modificar os dados armazenados no HBase cluster.

Note

O exemplo a seguir usa `protobuf-java-2.4.0a.jar` na versão AMI 2.3.3, mas você deve modificar o exemplo para corresponder à sua versão. Para verificar qual versão dos buffers de protocolo JAR você tem, execute o comando no prompt de comando do 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 ;
```

Personalizando a configuração HBase

Embora as configurações padrão devam funcionar para a maioria dos aplicativos, você tem a flexibilidade de modificá-las. HBase Para fazer isso, execute um dos dois scripts de ação de bootstrap:

- `configure-hbase-daemons`— Configura as propriedades dos daemons master, regionserver e zookeeper. Essas propriedades incluem o tamanho da pilha e as opções a serem passadas para a Java Virtual Machine (JVM) quando o HBase daemon é iniciado. Você pode definir essas propriedades como argumentos na ação de bootstrap. Essa ação de bootstrap modifica o arquivo de configuração `hbase-user-env /home/hadoop/conf/ .sh` no cluster. HBase
- `configure-hbase` — Define configurações HBase específicas do site, como a porta à qual o HBase mestre deve se vincular e o número máximo de vezes que o cliente deve repetir uma ação. CLI Você pode defini-los one-by-one como argumentos na ação de bootstrap ou especificar a localização de um arquivo de XML configuração no Amazon S3. Essa ação de bootstrap modifica o arquivo de configuração `/home/hadoop/conf/hbase-site.xml` no HBase cluster.

Note

Esses scripts, como outras ações de bootstrap, só podem ser executados quando o cluster é criado; você não pode usá-los para alterar a configuração de um HBase cluster em execução no momento.

Quando você executa as ações `configure-hbase` ou `configure-hbase-daemonsbootstrap`, os valores especificados substituem os valores padrão. Todos os valores não explicitamente definidos recebem os valores padrão.

Configurar HBase com essas ações de bootstrap é análogo ao uso de ações de bootstrap na Amazon para definir as configurações do Hadoop e as propriedades do EMR daemon do Hadoop. A diferença é que HBase não tem opções de memória por processo. Em vez disso, as opções de memória são definidas usando o `--daemon-opts` argumento, onde *daemon* é substituído pelo nome do daemon a ser configurado.

Configurar HBase daemons

EMRA Amazon fornece uma ação de bootstrap, `s3://region.elasticmapreduce/bootstrap-actions/configure-hbase-daemons`, que você pode usar para alterar a configuração dos HBase daemons, onde *region* é a região na qual você está lançando seu HBase cluster.

Para configurar HBase daemons usando o AWS CLI, adicione a ação `configure-hbase-daemons bootstrap` ao iniciar o cluster para configurar um ou mais daemons. HBase É possível definir as seguintes propriedades.

Propriedade	Descrição
<code>hbase-master-opts</code>	Opções que controlam como o daemon principal JVM é executado. Se definidas, elas substituem as OPTS variáveis <code>HBASE _ MASTER _</code> padrão.
<code>regionserver-opts</code>	Opções que controlam como o daemon do servidor da região JVM é executado. Se definidas, elas substituem as OPTS variáveis <code>HBASE _ REGIONSERVER _</code> padrão.
<code>zookeeper-opts</code>	Opções que controlam como o JVM daemon do zookeeper é executado. Se definidas, elas substituem as OPTS variáveis <code>HBASE _ ZOOKEEPER _</code> padrão.

Para obter mais informações sobre essas opções, consulte [hbase-env.sh](#) na HBase documentação.

O exemplo a seguir mostra uma ação de bootstrap para configurar valores para `zookeeper-opts` e `hbase-master-opts`.

Note

Os caracteres de continuação de linha do Linux (`\`) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (`^`).

```
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"]
```

Definir as configurações do HBase site

EMRA Amazon fornece uma ação de bootstraps3://elasticmapreduce/bootstrap-actions/configure-hbase,, que você pode usar para alterar a configuração do HBase. Você pode definir valores de configuração one-by-one, como argumentos na ação bootstrap, ou você pode especificar a localização de um arquivo de XML configuração no Amazon S3. Definir valores de configuração one-by-one é útil se você precisar definir apenas algumas configurações. Configurá-los usando um XML arquivo é útil se você tiver muitas alterações a fazer ou se quiser salvar suas configurações para reutilização.

Note

Você pode prefixar o nome do bucket do Amazon S3 com um prefixo de região, como, onde `s3://region.elasticmapreduce/bootstrap-actions/configure-hbase` *region* é a região na qual você está lançando seu HBase cluster.

Essa ação de bootstrap modifica o arquivo `/home/hadoop/conf/hbase-site.xml` de configuração no HBase cluster. A ação bootstrap só pode ser executada quando o HBase cluster é iniciado.

Para obter mais informações sobre as configurações do HBase site que você pode definir, consulte [Configuração padrão](#) na HBase documentação.

Defina a ação de `configure-hbase bootstrap` ao iniciar o HBase cluster e especifique os valores a `hbase-site.xml` serem alterados.

Para especificar configurações individuais HBase do site usando o AWS CLI

- Para alterar a `hbase.hregion.max.filesize` configuração, digite o seguinte comando e substitua *myKey* com o nome do seu par de EC2 chaves da Amazon.

Note

Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

```
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"]
```

Para especificar as configurações do HBase site com um XML arquivo usando o AWS CLI

1. Crie uma versão personalizada de `hbase-site.xml`. Seu arquivo personalizado deve ser válido XML. Para reduzir a chance de introduzir erros, comece com a cópia padrão de `hbase-site.xml`, localizada no nó EMR HBase principal da Amazon em `/home/hadoop/conf/hbase-site.xml`, e edite uma cópia desse arquivo em vez de criar um arquivo do zero. Você pode dar um novo nome para o seu novo arquivo ou manter `hbase-site.xml`.
2. Carregue seu arquivo `hbase-site.xml` personalizado em um bucket do Amazon S3. Ele deve ter permissões definidas para que a AWS conta que executa o cluster possa acessar o arquivo. Se a AWS conta que está lançando o cluster também possuir o bucket do Amazon S3, ela terá acesso.
3. Defina a ação de bootstrap `configure-hbase` ao iniciar o HBase cluster e inclua a localização do seu arquivo personalizado. `hbase-site.xml` O exemplo a seguir define os valores de configuração do HBase site para aqueles especificados no arquivos `s3://mybucket/my-hbase-site.xml`. Digite o seguinte comando, substitua `myKey` com o nome do seu EC2 key pair e substitua `mybucket` com o nome do seu bucket Amazon S3.

Note

Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

```
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=["--site-config-file", "s3://mybucket/config.xml"]
```

Se você especificar mais de uma opção para personalizar a HBase operação, deverá prefixar cada par de valores-chave com uma `-s` opção, conforme mostrado no exemplo a seguir:

```
--bootstrap-actions s3://elasticmapreduce/bootstrap-actions/configure-
hbase,Args=["-s", "zookeeper.session.timeout=60000"]
```

Com o proxy definido e a SSH conexão aberta, você pode visualizar a HBase interface do usuário abrindo uma janela do navegador com `http://master-public-dns-name:60010/master-status`, onde `master-public-dns-name` é o DNS endereço público do nó principal no HBase cluster.

Você pode visualizar os HBase registros atuais usando SSH para se conectar ao nó principal e navegar até o `mnt/var/log/hbase` diretório. Esses logs não estão disponíveis depois que o cluster é encerrado, a menos que você habilite o registro em log no Amazon S3 quando o cluster for iniciado.

Backup e restauração HBase

A Amazon EMR oferece a capacidade de fazer backup de seus HBase dados no Amazon S3, manualmente ou de forma automatizada. Você pode executar backups completos e incrementais. Depois de ter uma versão de backup dos HBase dados, você pode restaurar essa versão em um

HBase cluster. Você pode restaurar em um HBase cluster que esteja em execução no momento ou iniciar um novo cluster pré-preenchido com dados de backup.

Durante o processo de backup, HBase continua executando comandos de gravação. Embora isso garanta que o cluster permanecerá disponível durante o backup, existe o risco de inconsistência entre os dados que estão sendo copiados para backup e quaisquer operações de gravação que estejam sendo executadas em paralelo. Para entender as inconsistências que podem surgir, é preciso considerar que HBase distribui as operações de gravação entre os nós em seu cluster. Se uma operação de gravação acontecer depois que um nó específico for sondado, esses dados não serão incluídos no arquivamento de backup. Você pode até descobrir que gravações anteriores no HBase cluster (enviadas para um nó que já foi pesquisado) podem não estar no arquivo de backup, enquanto gravações posteriores (enviadas para um nó antes de ser pesquisado) estão incluídas.

Se for necessário um backup consistente, você deverá pausar as gravações HBase durante a parte inicial do processo de backup, a sincronização entre os nós. Isso pode ser feito especificando o parâmetro `--consistent` ao solicitar um backup. Com esse parâmetro, as gravações durante esse período são enfileiradas e executadas logo após a conclusão da sincronização. Você também pode agendar backups recorrentes, que resolvem quaisquer inconsistências com o passar do tempo, conforme os dados perdidos em uma passagem são copiados para backup na passagem seguinte.

Ao fazer backup HBase dos dados, você deve especificar um diretório de backup diferente para cada cluster. Uma maneira fácil de fazer isso é usar o identificador de cluster como parte do caminho especificado para o diretório de backup. Por exemplo, `s3://mybucket/backups/j-3AEXXXXXX16F2`. Isso garante que qualquer backup incremental futuro faça referência ao HBase cluster correto.

Quando você estiver pronto para excluir arquivos de backup antigos que não são mais necessários, recomendamos que você primeiro faça um backup completo dos seus HBase dados. Isso garante que todos os dados sejam preservados e fornece uma base para futuros backups incrementais. Terminado o backup completo, você pode navegar até o local do backup e excluir manualmente os arquivos de backup antigos.

O processo HBase de backup usa o S3 DistCp para a operação de cópia, que tem certas limitações em relação ao espaço de armazenamento temporário de arquivos.

Faça backup e restaure HBase usando o console

O console oferece a capacidade de iniciar um novo cluster e preenchê-lo com dados de um HBase backup anterior. Ele também oferece a capacidade de programar backups incrementais periódicos

dos HBase dados. Funcionalidades adicionais de backup e restauração, como a capacidade de restaurar dados em um cluster já em execução, fazer backups manuais e agendar backups completos automatizados, estão disponíveis usando CLI o.

Para preencher um novo cluster com HBase dados arquivados usando o console

1. Navegue até o novo EMR console da Amazon e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte [Usar o console antigo](#).
2. Selecione Criar cluster.
3. Na seção Configuração de software, para Aplicativos adicionais, escolha Configurar HBasee adicionar.
4. Na caixa de diálogo Add Application (Adicionar aplicativo), marque Restore From Backup (Restaurar a partir do backup).
5. Em Local do Backup, especifique o local do backup a ser carregado no novo HBase cluster. Esse deve ser um Amazon S3 URL do formulário. `s3://myawsbucket/backups/`
6. Para Backup Version (Versão de backup), você tem a opção de especificar o nome de uma versão de backup a ser carregada, definindo um valor. Se você não definir um valor para a versão de backup, a Amazon EMR carrega o backup mais recente no local especificado.
7. Escolha Add (Adicionar) e continue para criar o cluster com outras opções, conforme desejado.

Para agendar backups automatizados de HBase dados usando o console

1. Na seção Configuração de software, para Aplicativos adicionais, escolha Configurar HBasee adicionar.
2. Escolha Schedule Regular Backups (Programar backups regulares).
3. Especifica se os backups devem ser consistentes. Um backup consistente é aquele que pausa operações de gravação durante o estágio de backup inicial, a sincronização entre nós. Qualquer operação de gravação pausada é colocada em uma fila e retomada quando a sincronização é concluída.
4. Defina com que frequência os backups devem ocorrer, inserindo um número para Backup Frequency (Frequência de backup) e escolhendo Days (Dias), Hours (Horas) ou Minutes (Minutos). O primeiro backup automatizado executado é um backup completo; depois disso, a Amazon EMR salva backups incrementais com base na programação especificada por você.

5. Especifique o local no Amazon S3 em que os backups devem ser armazenados. O backup de cada HBase cluster deve ser feito em um local separado no Amazon S3 para garantir que os backups incrementais sejam calculados corretamente.
6. Especifique quando o primeiro backup deve ocorrer, definindo um valor para Backup Start Time (Hora de início do backup). Você pode definir isso comonow, o que faz com que o primeiro backup seja iniciado assim que o cluster estiver em execução ou inserir uma data e hora no [ISOformato](#). Por exemplo, 2013-09-26T 20:00 Z, define o horário de início para 26 de setembro de 2013 às 20:00. UTC
7. Escolha Adicionar.
8. Continue com a criação do cluster, com outras opções conforme desejado.

Monitor HBase com CloudWatch

A Amazon EMR relata três métricas CloudWatch que você pode usar para monitorar seus HBase backups. Essas métricas são enviadas CloudWatch em intervalos de cinco minutos e são fornecidas gratuitamente.

Métrica	Descrição
HBaseBackupFailed	<p>Se o último backup falhou. É definido como 0 por padrão e atualizado para 1 se a tentativa de backup anterior falhou. Essa métrica é relatada somente para HBase clusters.</p> <p>Caso de uso: monitorar HBase backups</p> <p>Unidades: Contagem</p>
HBaseMostRecentBackupDuration	<p>O tempo que o backup anterior levou para ser concluído . Essa métrica é definida independentemente de o último backup ter sido bem-sucedido ou ter falhado. Enquanto o backup está sendo processado, essa métrica retorna o número de minutos desde o início do backup. Essa métrica é relatada somente para HBase clusters.</p> <p>Caso de uso: Monitorar HBase backups</p>

Métrica	Descrição
	Unidade: minutos
HBaseTimeSinceLastSuccessfulBackup	<p>O número de minutos decorridos após o início do último HBase backup bem-sucedido em seu cluster. Essa métrica é relatada somente para HBase clusters.</p> <p>Caso de uso: monitorar HBase backups</p> <p>Unidade: minutos</p>

Configurar o Ganglia para HBase

Você configura o Ganglia para HBase usar a ação `configure-hbase-for-gangliabootstrap`. Essa ação de bootstrap é configurada HBase para publicar métricas no Ganglia.

Você deve configurar o HBase Ganglia ao iniciar o cluster; os relatórios do Ganglia não podem ser adicionados a um cluster em execução.

O Ganglia também armazena arquivos de log no servidor em `/mnt/var/log/ganglia/rrds`. Se você tiver configurado o cluster para manter a persistência dos arquivos de log em um bucket do Amazon S3, os arquivos de log do Ganglia serão mantidos lá também.

Para iniciar um cluster com o Ganglia forHBase, use a ação `configure-hbase-for-gangliabootstrap` conforme mostrado no exemplo a seguir.

Note

Os caracteres de continuação de linha do Linux (`\`) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (`^`).

```
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
```

Depois que o cluster é executado com o Ganglia configurado, você pode acessar os gráficos e relatórios do Ganglia usando a interface gráfica em execução no nó principal.

Especificações do aplicativo Pig para AMI versões anteriores da Amazon EMR

Versões compatíveis do Pig

A versão do Pig que você pode adicionar ao seu cluster depende da versão do Amazon EMR AMI e da versão do Hadoop que você está usando. A tabela abaixo mostra quais AMI versões e versões do Hadoop são compatíveis com as diferentes versões do Pig. Recomendamos que você use a versão mais recente do Pig para aproveitar as melhorias de desempenho e as novas funcionalidades.

Quando você usa o API para instalar o Pig, a versão padrão é usada, a menos que você especifique `--pig-versions` como argumento para a etapa que carrega o Pig no cluster durante a chamada para. [RunJobFlow](#)

Versão do Pig	AMI versão	Parâmetros de configuração	Detalhes da versão do Pig
0.12.0	3.1.0 e posterior	<code>--ami-version 3.1</code>	Adiciona suporte para:
Notas de release		<code>--ami-version 3.2</code>	<ul style="list-style-type: none"> Streaming UDFs sem JVM implementações
Documentação		<code>--ami-version 3.3</code>	<ul style="list-style-type: none"> ASSERTe operadores IN CASE expressão AvroStorage como uma função integrada do Pig. ParquetLoader e ParquetStorer como funções integradas

Versão do Pig	AMIversão	Parâmetros de configuração	Detalhes da versão do Pig
			<ul style="list-style-type: none"> • BigInteger e BigDecimal tipos
0.11.1.1 Notas de release Documentação	2.2 e posterior	<pre>--pig-versions 0.11.1.1 --ami-version 2.2</pre>	Melhora o desempenho do LOAD comando PigStorage se a entrada residir no Amazon S3.
0.11.1 Notas de release Documentação	2.2 e posterior	<pre>--pig-versions 0.11.1 --ami-version 2.2</pre>	Adiciona suporte para JDK 7, Hadoop 2, funções definidas pelo usuário do Groovy, SchemaTup le otimização, novos operadores e muito mais. Para obter mais informações, consulte Log de alterações do Pig 0.11.1 .
0.9.2.2 Notas de release Documentação	2.2 e posterior	<pre>--pig-versions 0.9.2.2 --ami-version 2.2</pre>	Adiciona suporte para Hadoop 1.0.3.
0.9.2.1 Notas de release Documentação	2.2 e posterior	<pre>--pig-versions 0.9.2.1 --ami-version 2.2</pre>	Adiciona suporte para MapR.

Versão do Pig	AMIversão	Parâmetros de configuração	Detalhes da versão do Pig
0.9.2 Notas de release Documentação	2.2 e posterior	<code>--pig-versions 0.9.2</code> <code>--ami-version 2.2</code>	Inclui várias melhorias no desempenho e correções de erros. Para obter informações completas sobre as alterações do Pig 0.9.2, acesse o Log de alterações do Pig 0.9.2 .
0.9.1 Notas de release Documentação	2,0	<code>--pig-versions 0.9.1</code> <code>--ami-version 2.0</code>	
0.6 Notas de release	1,0	<code>--pig-versions 0.6</code> <code>--ami-version 1.0</code>	
0.3 Notas de release	1,0	<code>--pig-versions 0.3</code> <code>--ami-version 1.0</code>	

Detalhes da versão do Pig

A Amazon EMR oferece suporte a determinadas versões do Pig que podem ter EMR patches adicionais da Amazon aplicados. Você pode configurar qual versão do Pig será executada nos EMR clusters da Amazon. Para obter mais informações sobre como fazer isso, consulte [Apache Pig](#). As seções a seguir descrevem as diferentes versões do Pig e os patches aplicados às versões carregadas na AmazonEMR.

Patches do Pig

Esta seção descreve os patches personalizados aplicados às versões do Pig disponíveis na AmazonEMR.

Patches do Pig 0.11.1.1

A EMR versão Amazon do Pig 0.11.1.1 é uma versão de manutenção que melhora o desempenho do LOAD comando PigStorage se a entrada residir no Amazon S3.

Patches do Pig 0.11.1

A EMR versão Amazon do Pig 0.11.1 contém todas as atualizações fornecidas pela Apache Software Foundation e os EMR patches cumulativos da Amazon do Pig versão 0.9.2.2. No entanto, não há novos patches EMR específicos da Amazon no Pig 0.11.1.

Patches do Pig 0.9.2

A Apache Pig 0.9.2 é uma versão de manutenção do Pig. A EMR equipe da Amazon aplicou os seguintes patches na EMR versão Amazon do Pig 0.9.2.

Patch	Descrição
PIG-1429	<p>Adiciona o tipo de dados Boolean ao Pig como um tipo de dados de primeira classe. Para obter mais informações, acesse https://issues.apache.org/jira/browse/ PIG -1429.</p> <p>Status: Confirmado</p> <p>Fixed in Apache Pig Version (Corrigido no Apache Pig versão): 0.10</p>
PIG-1824	<p>Support módulos de importação em JythonUDF. Para obter mais informações, acesse https://issues.apache.org/jira/browse/ PIG -1824.</p> <p>Status: Confirmado</p> <p>Fixed in Apache Pig Version (Corrigido no Apache Pig versão): 0.10</p>

Patch	Descrição
PIG-2010	<p>Pacote registrado JARs no cache distribuído. Para obter mais informações, acesse https://issues.apache.org/jira/browse/PIG-2010.</p> <p>Status: Confirmado</p> <p>Fixed in Apache Pig Version (Corrigido no Apache Pig versão): 0.11</p>
PIG-2456	<p>Adiciona um arquivo ~ /.pigbootstrap onde o usuário pode especificar instruções padrão do Pig. Para obter mais informações, acesse https://issues.apache.org/jira/browse/PIG-2456.</p> <p>Status: Confirmado</p> <p>Fixed in Apache Pig Version (Corrigido no Apache Pig versão): 0.11</p>
PIG-2623	<p>Support o uso de caminhos do Amazon S3 para se registrar. UDFs Para obter mais informações, acesse https://issues.apache.org/jira/browse/PIG-2623.</p> <p>Status: Confirmado</p> <p>Fixed in Apache Pig Version (Corrigido no Apache Pig versão): 0.10, 0.11</p>

Patches do Pig 0.9.1

A EMR equipe da Amazon aplicou os seguintes patches à EMR versão do Pig 0.9.1 da Amazon.

Patch	Descrição
JARArquivos de suporte e scripts Pig em dfs	<p>Adicione suporte para execução de scripts e registro de JAR arquivos armazenados no HDFS Amazon S3 ou em outros sistemas de arquivos distribuídos. Para obter mais informações, acesse https://issues.apache.org/jira/browse/PIG-1505.</p>

Patch	Descrição
	<p>Status: Confirmado</p> <p>Fixed in Apache Pig Version (Corrigido no Apache Pig versão) 0.8.0</p>
Oferece suporte a vários sistemas de arquivos no Pig	<p>Adiciona suporte a scripts do Pig para a leitura de dados de um sistema de arquivos e gravação em outro sistema. Para obter mais informações, acesse https://issues.apache.org/jira/browse/PIG-1564.</p> <p>Status: Não confirmado</p> <p>Fixed in Apache Pig Version (Corrigido no Apache Pig versão) n/a</p>
Adicionar data, hora e string do Piggybank UDFs	<p>Adicione data e hora e string UDFs para oferecer suporte a scripts personalizados do Pig. Para obter mais informações, acesse https://issues.apache.org/jira/browse/PIG-1565.</p> <p>Status: Não confirmado</p> <p>Fixed in Apache Pig Version (Corrigido no Apache Pig versão) n/a</p>

Clusters interativos e em lote no Pig

A Amazon EMR permite que você execute scripts do Pig em dois modos:

- Interativo
- Lote

Ao iniciar um cluster de longa execução usando o console ou o AWS CLI, você pode se conectar usando ssh o nó principal como usuário do Hadoop e usar o shell Grunt para desenvolver e executar seus scripts Pig de forma interativa. Usar o Pig no modo interativo permite que você revise o script do Pig com mais facilidade do que no modo em lote. Depois de revisar o script do Pig no modo interativo com êxito, você pode carregar o script para o Amazon S3 e executar o script no modo em

lote na produção. Você também pode enviar comandos do Pig interativamente em um cluster em execução para analisar e transformar os dados conforme necessário.

No modo em lote, você carrega o script do Pig para o Amazon S3 e, em seguida, envia o trabalho para o cluster como uma etapa. As etapas do Pig podem ser enviadas para um cluster de execução prolongada ou um cluster temporário.

Especificações do aplicativo Spark com AMI versões anteriores da Amazon EMR

Usar o Spark interativamente ou no modo em lote

A Amazon EMR permite que você execute aplicativos Spark em dois modos:

- Interativo
- Lote

Ao iniciar um cluster de longa execução usando o console ou o AWS CLI, você pode se conectar usando SSH o nó principal como usuário do Hadoop e usar o shell do Spark para desenvolver e executar seus aplicativos Spark de forma interativa. Usar o Spark interativamente permite criar protótipos ou testar aplicativos Spark com mais facilidade do que em um ambiente em lote. Depois de revisar com sucesso o aplicativo Spark no modo interativo, você pode colocar esse aplicativo ou programa JAR Python no sistema de arquivos local do nó principal do cluster no Amazon S3. Em seguida, você pode enviar o aplicativo como um fluxo de trabalho em lote.

No modo em lote, carregue o script do Spark no Amazon S3 ou no sistema de arquivos do nó principal local e, em seguida, envie o trabalho ao cluster como uma etapa. As etapas do Spark podem ser enviadas a um cluster de longa execução ou a um cluster transitório.

Criar um cluster com o Spark instalado

Para executar um cluster com o Spark instalado usando o console

1. Navegue até o novo EMR console da Amazon e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte [Usar o console antigo](#).
2. Selecione Criar cluster.
3. Para Configuração de software, escolha a versão de AMI lançamento que você precisa.

4. Para Applications to be installed (Aplicativos a serem instalados), escolha Spark na lista e depois escolha Configure and add (Configurar e adicionar).
5. Adicione argumentos para alterar a configuração do Spark conforme desejado. Para obter mais informações, consulte [Configurar o Spark](#). Escolha Adicionar.
6. Selecione outras opções conforme necessário e escolha Create cluster (Criar cluster).

O seguinte exemplo mostra como criar um cluster com o Spark usando 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);
```

Configurar o Spark

Você configura o Spark ao criar um cluster executando a ação bootstrap localizada no repositório [aws-labs/emr-bootstrap-actions/spark](#) no Github. Para argumentos que a ação bootstrap aceita, consulte o [README](#) nesse repositório. A ação de bootstrap configura propriedades no arquivo `$SPARK_CONF_DIR/spark-defaults.conf`. Para obter mais informações sobre configurações, consulte o tópico Configuração do Spark na documentação do Spark. Você pode substituir “mais recente” a seguir URL pelo número da versão do Spark que você está instalando, por exemplo, `2.2.0` <http://spark.apache.org/docs/latest/configuration.html>.

Você também pode configurar o Spark dinamicamente no momento do envio de cada aplicativo. Uma configuração para maximizar automaticamente a alocação de recursos para um executor está disponível usando o arquivo de configuração `spark`. Para obter mais informações, consulte [Substituir as definições de configuração padrão do Spark](#).

Alterar as configurações padrão do Spark

O seguinte exemplo mostra como criar um cluster com o `spark.executor.memory` definido como 2G usando a AWS CLI.

Note

Os caracteres de continuação de linha do Linux (`\`) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (`^`).

```
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
```

Enviar trabalhos para o Spark

Para enviar trabalhos para um cluster, use uma etapa para executar o `spark-submit` script em seu EMR cluster. Adicione a etapa usando o `addJobFlowSteps` método em [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);
```

Substituir as definições de configuração padrão do Spark

Você pode querer substituir valores de configuração padrão do Spark para cada aplicativo. Você pode fazer isso ao enviar aplicativos usando uma etapa, que transmite essencialmente opções para `spark-submit`. Por exemplo, você pode querer alterar a memória alocada para um processo de executor modificando `spark.executor.memory`. Você pode fornecer ao switch `--executor-memory` um argumento semelhante ao seguinte:

```
/home/hadoop/spark/bin/spark-submit --executor-memory 1g --class
org.apache.spark.examples.SparkPi /home/hadoop/spark/lib/spark-examples*.jar 10
```

Da mesma forma, você pode ajustar `--executor-cores` e `--driver-memory`. Em uma etapa, você forneceria os seguintes argumentos para a etapa:

```
--executor-memory 1g --class org.apache.spark.examples.SparkPi /home/hadoop/spark/lib/
spark-examples*.jar 10
```

Você também pode ajustar as configurações que não possam não ter uma opção interna usando `--conf`. Para obter mais informações sobre outras configurações que são ajustáveis, consulte o tópico [Dynamically loading Spark properties](#) na documentação do Apache Spark.

Diferenças de DistCp utilitários do S3 com AMI versões anteriores da Amazon EMR

DistCp Versões S3 suportadas na Amazon EMR

As seguintes DistCp versões do S3 são suportadas nas EMR AMI versões da Amazon. DistCpAs versões S3 posteriores à 1.0.7 são encontradas diretamente nos clusters. Use o JAR in `/home/hadoop/lib` para obter os recursos mais recentes.

Version (Versão)	Descrição	Data de lançamento
1.0.8	Adiciona as opções <code>--appendToLastFile</code> , <code>--requirePreviousManifest</code> e <code>--storageClass</code> .	3 de janeiro de 2014
1.0.7	Adiciona a opção <code>--s3ServerSideEncryption</code> .	2 de maio de 2013

Version (Versão)	Descrição	Data de lançamento
1.0.6	Adiciona a opção <code>--s3Endpoint</code> .	6 de agosto de 2012
1.0.5	Melhora a capacidade de especificar qual versão do S3 DistCp deve ser executada.	27 de junho de 2012
1.0.4	Melhora a opção <code>--deleteOnSuccess</code> .	19 de junho de 2012
1.0.3	Adiciona suporte às opções <code>--numberOfFiles</code> e <code>--startingIndex</code> .	12 de junho de 2012
1.0.2	Melhora a nomeação de arquivos ao usar grupos.	6 de junho de 2012
1.0.1	Lançamento inicial do S3DistCp.	19 de janeiro de 2012

Adicionar uma etapa de DistCp cópia do S3 a um cluster

Para adicionar uma etapa de DistCp cópia do S3 a um cluster em execução, digite o seguinte comando, substitua `j-3GYXXXXXX9I0K` com seu ID de cluster e substitua `mybucket` com o nome do seu bucket do Amazon S3.

Note

Os caracteres de continuação de linha do Linux (`\`) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (`^`).

```
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,]+"
```

Exemplo Carregue CloudFront os logins da Amazon em HDFS

Este exemplo carrega CloudFront os logins da Amazon HDFS adicionando uma etapa a um cluster em execução. No processo, ele altera o formato de compressão de Gzip (o CloudFront padrão) para LZO. Isso é útil porque os dados compactados usando LZO podem ser divididos em vários mapas à medida que são descompactados, então você não precisa esperar até que a compactação seja concluída, como acontece com o Gzip. Isso proporciona um melhor desempenho quando você analisa os dados usando a AmazonEMR. Esse exemplo também melhora a performance ao usar a expressão regular especificada na opção `--groupBy` para combinar todos os logs de uma determinada hora em um único arquivo. Os EMR clusters da Amazon são mais eficientes ao processar alguns arquivos grandes e LZO compactados do que ao processar muitos arquivos pequenos compactados com GZIP. Para dividir LZO arquivos, você deve indexá-los e usar a biblioteca de terceiros `hadoop-lzo`.

Para carregar CloudFront os logins da AmazonHDFS, digite o seguinte comando, substitua `j-3GYXXXXXX9I0K` com seu ID de cluster e substitua `mybucket` com o nome do seu bucket do Amazon S3.

Note

Os caracteres de continuação de linha do Linux (`\`) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (`^`).

```
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"]
```

Considere o caso em que o exemplo anterior é executado nos seguintes arquivos de CloudFront log.

```
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
```

O S3 DistCp copia, concatena e compacta os arquivos nos dois arquivos a seguir, onde o nome do arquivo é determinado pela correspondência feita pela expressão regular.

```
hdfs:///local/2012-02-23-01.lzo  
hdfs:///local/2012-02-23-02.lzo
```

O que há de novo?

Esta página descreve as mudanças e funcionalidades disponíveis nas versões mais recentes da Amazon EMR 7.x, 6.x e 5.x.

Essas notas de lançamento também estão disponíveis nas páginas [Amazon EMR 7.2.0](#), [Amazon EMR 6.15.0](#) e [Amazon EMR 5.36.2](#), junto com as versões do aplicativo, as versões dos componentes e as classificações de configuração disponíveis para cada versão.

- Para obter notas de versões anteriores, consulte o [EMR Arquivo de notas de lançamento da Amazon](#).
- Para receber atualizações quando um novo EMR lançamento da Amazon estiver disponível, assine o [RSSfeed para ver as notas de EMR lançamento da Amazon](#).

Note

Versões posteriores da Amazon EMR usam o AWS Signature Version 4 (SigV4) para autenticar solicitações no Amazon S3. Recomendamos que você use uma EMR versão da Amazon compatível SigV4 para que você possa acessar novos buckets do S3 e evitar a interrupção de suas cargas de trabalho. Para obter mais informações e uma lista dos EMR lançamentos da Amazon compatíveis SigV4, consulte [Amazon EMR e AWS Signature versão 4](#).

Amazon EMR 7.2.0 (versão mais recente da série 7.x)

Novos EMR lançamentos da Amazon são disponibilizados em diferentes regiões por um período de vários dias, começando com a primeira região na data de lançamento inicial. A versão mais recente pode não estar disponível em sua região durante esse período.

As notas de lançamento a seguir incluem informações sobre a EMR versão 7.2.0 da Amazon. As alterações são relativas à versão 7.2.0.

Novos atributos

- Atualizações da aplicação: Amazon EMR 7.2.0 application upgrades include Iceberg 1.5.0-amzn-0 and Delta 3.1.0.

- EMRA Amazon adiciona suporte para que você possa usar outros aplicativos HBase, como Flink e Hive, com a classe de armazenamento Amazon S3 Express One Zone.
- Esta versão adiciona a capacidade de ler objetos restaurados, para que você possa ler objetos Glacier de um local do S3 com o S3A protocolo. Esse recurso funciona com Spark, Flink e Hive.
- Rótulos de nós com escalabilidade gerenciada - Use rótulos de nós com suas instâncias com base no tipo de mercado ou tipo de nó para melhorar a escalabilidade automática da Amazon. EMR Para obter mais informações, consulte [Usando escalabilidade gerenciada na Amazon EMR](#).

Problemas conhecidos

- O Python 3.11 não é compatível com o Studio. EMR

Alterações, melhorias e problemas resolvidos

- Esta versão corrige um problema de impasse que pode ocorrer durante as operações internas de limpeza de etapas. Essa operação gerencia o ciclo de vida das etapas à medida que elas são concluídas no EMR cluster. Esse problema afeta EMR operações críticas da Amazon, como operação por etapas e escalabilidade.
- Esta versão resolve um problema em que clusters personalizados com determinados arquivos de log preexistentes podem fazer com AMIs que o daemon de gerenciamento de EMR log da Amazon falhe.
- O Amazon EMR 7.2.0 atualiza o EMR daemon da Amazon responsável pelas atividades de gerenciamento e monitoramento de clusters da v1 para a v2. AWS SDK
- Quando você inicia um cluster com a versão de patch mais recente do Amazon EMR 5.36 ou superior, 6.6 ou superior ou 7.0 ou superior, a Amazon EMR usa a versão mais recente do Amazon Linux 2023 ou Amazon Linux 2 para a Amazon padrão. EMR AMI Para obter mais informações, consulte [Usando o Amazon Linux padrão AMI para a Amazon EMR](#).

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2023.5.2 240708.0	6.1.96-102.177.amzn2023	8 de julho de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv), Oeste do Canadá (Calgary), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia)

Amazon EMR 6.15.0 (versão mais recente da série 6.x)

Novos EMR lançamentos da Amazon são disponibilizados em diferentes regiões por um período de vários dias, começando com a primeira região na data de lançamento inicial. A versão mais recente pode não estar disponível em sua região durante esse período.

As notas de lançamento a seguir incluem informações sobre a EMR versão 6.15.0 da Amazon. As alterações são referentes à versão 6.14.0. Para obter informações sobre o cronograma da versão, consulte o [Log de alterações 6.15.0](#).

Novos atributos

- Atualizações da aplicação: 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.
- [Lançamentos mais rápidos para EMR clusters executados em EC2](#) — Agora é até 35% mais rápido lançar um Amazon EMR on EC2 cluster. Com essa melhoria, a maioria dos clientes pode lançar seus clusters em cinco minutos ou menos.
- [CodeWhisperer para EMR Studio](#) — Agora você pode usar a Amazon CodeWhisperer com o Amazon EMR Studio para obter recomendações em tempo real enquanto escreve código JupyterLab. CodeWhisperer pode concluir seus comentários, concluir linhas únicas de código, fazer line-by-line recomendações e gerar funções totalmente formadas.
- [Tempos mais rápidos de reinício do trabalho com o Flink](#) — Com o Amazon EMR 6.15.0 e superior, vários novos mecanismos estão disponíveis para o Apache Flink para melhorar o tempo

de reinício do trabalho durante as operações de recuperação ou escalabilidade de tarefas. Isso otimiza a velocidade de recuperação e reinicialização dos gráficos de execução para melhorar a estabilidade do trabalho.

- [Controle de acesso detalhado e em nível de tabela para formatos de tabela aberta — Com o EMR Amazon 6.15.0 e superior, quando você executa trabalhos do Spark na Amazon em EC2 clusters que acessam dados EMR no AWS Glue Data Catalog, você pode AWS Lake Formation usar para aplicar permissões em nível de tabela, linha, coluna e célula em tabelas baseadas em Hudi, Iceberg ou Delta Lake.](#)
- **Atualização do Hadoop** — O Amazon EMR 6.15.0 inclui uma atualização do Apache Hadoop para a versão 3.3.6. O Hadoop 3.3.6 era a versão mais recente na época da implantação do Amazon EMR 6.15, lançada pela Apache em junho de 2023. Versões anteriores da Amazon EMR (6.9.0 a 6.14.x) usavam o Hadoop 3.3.3.

A atualização inclui centenas de melhorias e correções, além de recursos que incluem parâmetros de nódulos de dados reconfiguráveis, DFSAdmin opção de iniciar operações de reconfiguração em massa em todos os datanodes ativos e um vetor que permite API que leitores exigentes especifiquem vários intervalos para leitura. O Hadoop 3.3.6 também adiciona suporte HDFS APIs e semântica ao seu registro de gravação antecipada (WAL), para que ele HBase possa ser executado em outras implementações de sistemas de armazenamento. Para obter mais informações, consulte os logs de alterações das versões [3.3.4](#), [3.3.5](#) e [3.3.6](#) na documentação do Apache Hadoop.

- **Support AWS SDK for Java, versão 2** - Os aplicativos Amazon EMR 6.15.0 podem ser usados AWS SDK para Java nas versões [1.12.569](#) ou [2.20.160](#) se o aplicativo oferecer suporte à v2. O AWS SDK for Java 2.x é uma grande reescrita da base de código da versão 1.x. Ele foi criado com base no Java 8+ e adiciona vários recursos frequentemente solicitados. Isso inclui suporte para E/S sem bloqueio e a capacidade de conectar uma HTTP implementação diferente em tempo de execução. Para obter mais informações, incluindo um Guia de migração de SDK Java v1 para v2, consulte o guia [AWS SDK para Java, versão 2](#).

Alterações, melhorias e problemas resolvidos

- Para melhorar seus EMR clusters de alta disponibilidade, esta versão permite a conectividade com EMR daemons da Amazon em um host local que usa endpoints. IPv6
- Esta versão permite a comunicação TLS 1.2 com ZooKeeper provisionados em todos os nós primários do seu cluster de alta disponibilidade.

- Esta versão melhora o gerenciamento dos arquivos de log de ZooKeeper transações que são mantidos nos nós primários para minimizar os cenários em que os arquivos de log ultrapassam os limites e interrompem as operações do cluster.
- Essa versão torna a comunicação entre nós mais resiliente para clusters de alta disponibilidade. Essa melhoria reduz a chance de falhas na ação de bootstrap ou na inicialização do cluster.
- O Tez no Amazon EMR 6.15.0 introduz configurações que você pode especificar para abrir de forma assíncrona as divisões de entrada em uma divisão agrupada Tez. Isso resulta em uma performance mais rápida das consultas de leitura quando há um grande número de divisões de entrada em uma única divisão agrupada do Tez. Para obter mais informações, consulte a [Abertura assíncrona da divisão do Tez](#).
- Quando você inicia um cluster com a versão de patch mais recente do Amazon EMR 5.36 ou superior, 6.6 ou superior ou 7.0 ou superior, a Amazon EMR usa a versão mais recente do Amazon Linux 2023 ou Amazon Linux 2 para a Amazon padrão. Para obter mais informações, consulte [Usando o Amazon Linux padrão AMI para a Amazon EMR](#).

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 709.1	4.14.348	23 de julho de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			<p>Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), AWS GovCloud China (Pequim), China AWS GovCloud (Ningxia), Ásia Pacífico (Hyderabad), Oriente Médio (), Europa (EspanhaUAE), Europa (Zurique), Ásia-Pacífico (Melbourne), Israel (Tel Aviv), Oeste do Canadá (Calgary)</p>

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 223.0	4.14.336	8 de março de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia) , Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 131.0	4.14.336	14 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia) , Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 124.0	4.14.336	7 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelease Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia) , Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 109.0	4.14.334	24 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia) , Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 218.0	4.14.330	2 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 206.0	4.14.330	22 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelease Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 116.0	4.14.328	11 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 101.0	4.14.327	13 de novembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

Amazon EMR 5.36.2 (versão mais recente da série 5.x)

Novos EMR lançamentos da Amazon são disponibilizados em diferentes regiões por um período de vários dias, começando com a primeira região na data de lançamento inicial. A versão mais recente pode não estar disponível em sua região durante esse período.

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.36.2 da Amazon. As alterações são relativas ao 5.36.1. Para obter informações sobre o cronograma da versão, consulte o [log de alterações](#).

Alterações, melhorias e problemas resolvidos

- Esta versão melhora a lógica de redução do cluster para que a Amazon EMR não reduza os nós principais abaixo da configuração do fator de HDFS replicação do cluster. Essa melhoria atende aos requisitos de redundância de dados e reduz a chance de uma operação de escalabilidade parar.
- Esta versão adiciona um novo mecanismo de repetição ao fluxo de trabalho de escalabilidade de cluster para aquela execução Presto ou Trino. Essa melhoria reduz o risco de o redimensionamento do cluster ser executado indefinidamente devido a uma única falha na operação de redimensionamento. Ela também aprimora a utilização dos clusters, porque seu cluster aumenta e reduz a escala verticalmente com mais rapidez.
- Corrige um problema em que as operações de redução de escala do cluster podem parar enquanto a Amazon EMR descomissiona normalmente um nó principal e ele fica insalubre antes de ser totalmente desativado.

- Melhora a estabilidade de um nó em um cluster de alta disponibilidade com vários nós primários quando a Amazon EMR reinicia um único nó.
- Otimiza o gerenciamento de registros com a Amazon em EMR execução na AmazonEC2. Como resultado, é possível ver uma pequena redução nos custos de armazenamento dos logs do cluster.
- Melhora o gerenciamento dos arquivos de log de ZooKeeper transações que são mantidos nos nós primários para minimizar os cenários em que os arquivos de log ultrapassam os limites e interrompem as operações do cluster.
- Corrige um bug raro que pode fazer com que um cluster de alta disponibilidade com vários nós primários falhe por não conseguir se comunicar com o ResourceManager Yarn.
- Quando você inicia um cluster com a versão de patch mais recente do Amazon EMR 5.36 ou superior, 6.6 ou superior ou 7.0 ou superior, a Amazon EMR usa a versão mais recente do Amazon Linux 2023 ou Amazon Linux 2 para a Amazon padrão. EMR AMI Para obter mais informações, consulte [Usando o Amazon Linux padrão AMI para a Amazon EMR](#).

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 709.1	4.14.348	23 de julho de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			<p>Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), AWS GovCloud China (Pequim), China AWS GovCloud (Ningxia), Ásia Pacífico (Hyderabad), Oriente Médio (), Europa (EspanhaUAE), Europa (Zurique), Ásia-Pacífico (Melbourne), Israel (Tel Aviv), Oeste do Canadá (Calgary)</p>

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 503.0	4.14.343	xxxxxx, 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

Amazon EMR e AWS Signature versão 4

As EMR versões da Amazon usam o AWS Signature Version 4 (SigV4) para autenticar solicitações no Amazon S3. Os buckets criados no Amazon S3 depois de 24 de junho de 2020 não são compatíveis com solicitações assinadas pelo Signature Version 2 (SigV2). Os buckets criados em ou antes de 24 de junho de 2020 continuarão a oferecer suporte ao SigV2. Recomendamos que você migre para uma EMR versão da Amazon compatível para que você SigV4 possa acessar novos buckets do S3 e evitar a interrupção de suas cargas de trabalho.

Se você usa aplicativos incluídos na Amazon, EMR como Apache Spark, Apache Hive e Presto, não precisa alterar o código do aplicativo para usar SigV4. Se você usa aplicativos personalizados que não estão incluídos na AmazonEMR, talvez seja necessário atualizar seu código para usá-lo SigV4. Para obter mais informações, consulte [Migração do Signature Version 2 para o Signature Version 4](#) no Guia do usuário do Amazon S3.

Os seguintes EMR lançamentos da Amazon oferecem suporte 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. Todas as versões 6.x e 7.x são compatíveis com SigV4.

Abordagem para CVE mitigar -2021-44228

Note

Para a Amazon EMR versão 6.9.0 e posterior, todos os componentes instalados pela Amazon EMR que usam bibliotecas Log4j usam Log4j versão 2.17.1 ou posterior.

Amazon EMR rodando em EC2

O problema discutido em [CVE-2021-44228](#) é relevante para as versões principais do Apache Log4j entre 2.0.0 e 2.14.1 ao processar entradas de fontes não confiáveis. EMROs clusters da Amazon lançados com as versões EMR 5.x até 5.34.0 e EMR 6.x até o Amazon EMR 6.5.0 incluem estruturas de código aberto, como Apache Hive, Flink, Presto e Trino, que usam essas versões do Apache Log4j. HUDI No entanto, muitos clientes usam as estruturas de código aberto instaladas em seus EMR clusters da Amazon para processar e registrar entradas de fontes não confiáveis.

Recomendamos que você aplique a “Amazon EMR Bootstrap Action Solution for Log4j CVE -2021-44228” conforme descrito na seção a seguir. Essa solução também aborda CVE -2021-45046.

Note

Os scripts de ação de bootstrap para a Amazon EMR foram atualizados em 7 de setembro de 2022 para incluir correções incrementais de bugs e melhorias para o Oozie. Se você usa o Oozie, deve aplicar a solução de ação EMR bootstrap atualizada da Amazon descrita na seção a seguir.

Amazon EMR em EKS

Se você usa o [Amazon EMR on EKS](#) com a configuração padrão, não é afetado pelo problema descrito em CVE -2021-44228 e não precisa aplicar a solução descrita na seção. [Solução de ação EMR bootstrap da Amazon para Log4j -2021-44228 e -2021-45046 CVE CVE](#) Para o Amazon EMR onEKS, o Amazon EMR Runtime para Spark usa o Apache Log4j versão 1.2.17. Ao usar o Amazon EMR on, EKS você não deve alterar a configuração padrão `log4j . appender` do componente `paralog`.

Solução de ação EMR bootstrap da Amazon para Log4j -2021-44228 e -2021-45046 CVE CVE

Essa solução fornece uma ação de EMR bootstrap da Amazon que deve ser aplicada em seus EMR clusters da Amazon. Para cada EMR lançamento da Amazon, você encontrará um link para um script de ação de bootstrap abaixo. Para aplicar essa ação de bootstrap, você deve concluir as seguintes etapas:

1. Copie o script que corresponde à sua EMR versão da Amazon para um bucket S3 local em seu Conta da AWS. Verifique se você está usando um script de bootstrap específico para sua EMR versão da Amazon.
2. [Configure uma ação de bootstrap para que seus EMR clusters executem o script copiado para o bucket do S3 de acordo com as instruções descritas na documentação. EMR](#) Se você tiver outras ações de bootstrap configuradas para seus EMR clusters, certifique-se de que esse script esteja configurado como o primeiro script de ação de bootstrap a ser executado.
3. Encerre os EMR clusters existentes e inicie novos clusters com o script de ação bootstrap. AWS recomenda que você teste os scripts de bootstrap em seu ambiente de teste e valide seus aplicativos antes de aplicá-los ao seu ambiente de produção. Se você não estiver usando a

revisão mais recente para uma versão EMR secundária (por exemplo, 6.3.0), deverá usar a revisão mais recente (por exemplo, 6.3.1) e, em seguida, aplicar a solução discutida acima.

CVE-2021-44228 e -2021-45046 - Scripts de bootstrap para CVE lançamentos da Amazon EMR

Número de EMR lançamento da Amazon	Local do script	Data da versão do script
6.5.0	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-6.5.0-v2.sh</code>	24 de março de 2022
6.4.0	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-6.4.0-v2.sh</code>	24 de março de 2022
6.3.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-6.3.1-v2.sh</code>	24 de março de 2022
6.2.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-6.2.1-v2.sh</code>	24 de março de 2022
6.1.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-6.1.1-v2.sh</code>	14 de dezembro de 2021
6.0.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-6.0.1-v2.sh</code>	14 de dezembro de 2021

Número de EMR lançamento da Amazon	Local do script	Data da versão do script
5.34.0	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.34.0-v2.sh</code>	12 de dezembro de 2021
5.3.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.33.1-v2.sh</code>	12 de dezembro de 2021
5.32.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.32.1-v2.sh</code>	13 de dezembro de 2021
5.31.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.31.1-v2.sh</code>	13 de dezembro de 2021
5.30.2	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.30.2-v2.sh</code>	14 de dezembro de 2021
5.29.0	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.29.0-v2.sh</code>	14 de dezembro de 2021

Número de EMR lançamento da Amazon	Local do script	Data da versão do script
5.28.1	<pre>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.28.1-v2.sh</pre>	15 de dezembro de 2021
5.27.1	<pre>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.27.1-v2.sh</pre>	15 de dezembro de 2021
5.26.0	<pre>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.26.0-v2.sh</pre>	15 de dezembro de 2021
5.25.0	<pre>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.25.0-v2.sh</pre>	15 de dezembro de 2021
5.24.1	<pre>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.24.1-v2.sh</pre>	15 de dezembro de 2021
5.23.1	<pre>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.23.1-v2.sh</pre>	15 de dezembro de 2021

Número de EMR lançamento da Amazon	Local do script	Data da versão do script
5.22.0	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.22.0-v2.sh</code>	15 de dezembro de 2021
5.21.2	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.21.2-v2.sh</code>	15 de dezembro de 2021
5.20.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.20.1-v2.sh</code>	15 de dezembro de 2021
5.19.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.19.1-v2.sh</code>	15 de dezembro de 2021
5.18.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.18.1-v2.sh</code>	15 de dezembro de 2021
5.17.2	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.17.2-v2.sh</code>	15 de dezembro de 2021

Número de EMR lançamento da Amazon	Local do script	Data da versão do script
5.16.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.16.1-v2.sh</code>	15 de dezembro de 2021
5.15.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.15.1-v2.sh</code>	15 de dezembro de 2021
5.14.2	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.14.2-v2.sh</code>	15 de dezembro de 2021
5.13.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.13.1-v2.sh</code>	15 de dezembro de 2021
5.12.3	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.12.3-v2.sh</code>	15 de dezembro de 2021
5.11.4	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.11.4-v2.sh</code>	15 de dezembro de 2021

Número de EMR lançamento da Amazon	Local do script	Data da versão do script
5.10.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.10.1-v2.sh</code>	15 de dezembro de 2021
5.9.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.9.1-v2.sh</code>	15 de dezembro de 2021
5.8.3	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.8.3-v2.sh</code>	15 de dezembro de 2021
5.7.1	<code>s3://elasticmapreduce/ bootstrap-actions/ log4j/patch-log4j- emr-5.7.1-v2.sh</code>	15 de dezembro de 2021

EMR versão de lançamento	Revisão mais recente em dezembro de 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	5.3.1
5.32.0	5.32.1

EMR versão de lançamento	Revisão mais recente em dezembro de 2021
5.31.0	5.31.1
5.30.0 ou 5.30.1	5.30.2
5.28.0	5.28.1
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

Perguntas frequentes

- Os EMR lançamentos anteriores a EMR 5 são afetados por CVE -2021-44228?

Não. EMR versões anteriores à EMR versão 5 usam versões do Log4j anteriores à 2.0.

- Essa solução aborda CVE -2021-45046?

Sim, essa solução também aborda CVE-2021-45046.

- A solução lida com aplicativos personalizados que eu instalo em meus EMR clusters?

O script de bootstrap atualiza somente JAR os arquivos que são instalados pelo EMR. Se você instalar e executar aplicativos e JAR arquivos personalizados em seus EMR clusters por meio de ações de bootstrap, como etapas enviadas aos seus clusters, usando o Amazon Linux AMI personalizado ou por meio de qualquer outro mecanismo, trabalhe com o fornecedor do aplicativo para determinar se seus aplicativos personalizados são afetados pelo CVE -2021-44228 e determine uma solução apropriada.

- Como devo lidar com [imagens docker personalizadas com EMR on?](#) EKS

Se você adicionar aplicativos personalizados à EMR Amazon EKS usando [imagens docker personalizadas](#) ou enviar trabalhos para a Amazon EMR em arquivos de aplicativos EKS with personalizados, fale com o fornecedor do aplicativo para determinar se seus aplicativos personalizados são afetados pelo CVE -2021-44228 e determine uma solução apropriada.

- Como o script de bootstrap funciona para mitigar o problema descrito em CVE -2021-44228 e -2021-45046? CVE

O script bootstrap atualiza as instruções de EMR inicialização adicionando um novo conjunto de instruções. Essas novas instruções excluem os arquivos de JndiLookup classe usados por meio do Log4j por todas as estruturas de código aberto instaladas pelo. EMR Isso segue a [recomendação publicada pela Apache](#) para lidar com os problemas do Log4j.

- Existe uma atualização EMR que usa as versões 2.17.1 ou superiores do Log4j?

EMR5 versões até a versão 5.34 e EMR 6 versões até a versão 6.5 usam versões mais antigas de estruturas de código aberto que são incompatíveis com as versões mais recentes do Log4j. Se você continuar usando essas versões, recomendamos que você aplique a ação bootstrap para mitigar os problemas discutidos no. CVEs Após a versão EMR 5 5.34 e a versão EMR 6 6.5, os aplicativos que usam o Log4j 1.x e o Log4j 2.x serão atualizados para usar o Log4j 1.2.17

(ou superior) e o Log4j 2.17.1 (ou superior), respectivamente, e não exigirão o uso das ações de bootstrap fornecidas acima para mitigar os problemas. CVE

- Os EMR lançamentos são afetados pelo CVE -2021-45105?

Os aplicativos instalados pela Amazon EMR com EMR as configurações padrão não são afetados por CVE -2021-45105. Entre os aplicativos instalados pela AmazonEMR, somente o Apache Hive usa o Apache Log4j com [pesquisas de contexto](#) e não usa layout de padrão não padrão de forma a permitir que dados de entrada inadequados sejam processados.

- A Amazon é EMR afetada por alguma das CVE divulgações a seguir?

A tabela a seguir contém uma lista dos CVEs que estão relacionados ao Log4j e indica se cada um afeta a CVE Amazon. EMR As informações nesta tabela só se aplicam quando os aplicativos são instalados pela Amazon EMR usando as configurações padrão.

CVE	Impactos EMR	Observações
CVE-2022-23302	Não	A Amazon EMR não configura o Log4j JMSSink
CVE-2022-23305	Não	A Amazon EMR não configura o Log4j JDBCAppender
CVE-2022-23307	Não	A Amazon EMR não configura o Log4j Chainsaw
CVE-2020-9493	Não	A Amazon EMR não configura o Log4j Chainsaw
CVE-2021-44832	Não	EMRA Amazon não configura o Log4j JDBCAppender com uma string de conexão JNDI
CVE-2021-4104	Não	A Amazon EMR não usa o Log4j JMSAppender

CVE	Impactos EMR	Observações
CVE-2020-948	Não	Os aplicativos instalados pela Amazon EMR não usam o Log4j SMTPAppender
CVE-2019-17 571	Não	A Amazon EMR bloqueia o acesso público aos clusters e não inicia SocketServer
CVE-2019-17 531	Não	Recomendamos que você atualize para a EMR versão mais recente da Amazon. O Amazon EMR 5.33.0 e versões posteriores usam jackson-databind 2.6.7.4 ou posterior, e 6.1.0 e versões posteriores usam jackson-databind 2.10.0 ou posterior . EMR Essas versões do jackson-databind não são afetadas pelo. CVE

EMR Arquivo de notas de lançamento da Amazon

As notas de lançamento de todos os EMR lançamentos da Amazon estão disponíveis abaixo. Para obter informações de versão abrangentes para cada versão, consulte [Versões de lançamento do Amazon EMR 6.x](#), [Versões de lançamento do Amazon EMR 5.x](#) e [Versões de lançamento do Amazon EMR 4.x](#).

Para receber atualizações quando um novo EMR lançamento da Amazon estiver disponível, assine o [RSSfeed para ver as notas de EMR lançamento da Amazon](#).

Versão 6.14.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 6.14.0 da Amazon. As alterações são referentes à versão 6.13.0. Para obter informações sobre o cronograma da versão, consulte o [Log de alterações 6.14.0](#).

Novos atributos

- 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.
- A [escalabilidade EMR gerenciada da Amazon](#) agora está disponível na região ap-southeast-3 Ásia-Pacífico (Jacarta) para clusters que você cria com o Amazon EMR 6.14.0 ou superior.

Alterações, melhorias e problemas resolvidos

- A versão 6.14.0 otimiza o gerenciamento de registros com a Amazon em execução EMR na Amazon. EC2 Como resultado, é possível ver uma pequena redução nos custos de armazenamento dos logs do cluster.
- A versão 6.14.0 melhora o fluxo de trabalho de escalabilidade para considerar diferentes instâncias principais que têm uma variação substancial no tamanho de seus volumes da Amazon. EBS Essa melhoria se aplica somente aos nós centrais; as operações de redução dos nós de tarefas não são afetadas.
- A versão 6.14.0 melhora a forma como a Amazon EMR interage com aplicativos de código aberto, como. Apache Hadoop YARN ResourceManager and HDFS NameNode Essa melhoria reduz o risco de atrasos operacionais com o escalonamento do cluster e atenua as falhas de inicialização que ocorrem devido a problemas de conectividade com os aplicações de código aberto.
- A versão 6.14.0 otimiza a instalação da aplicação na inicialização do cluster. Isso melhora os tempos de inicialização do cluster para determinadas combinações de EMR aplicativos da Amazon.
- A versão 6.14.0 corrige um problema em que as operações de redução de escala do cluster podem parar quando um cluster executado em um domínio personalizado é reiniciado VPC no núcleo ou no nó da tarefa.
- Quando você inicia um cluster com a versão de patch mais recente do Amazon EMR 5.36 ou superior, 6.6 ou superior ou 7.0 ou superior, a Amazon EMR usa a versão mais recente do Amazon Linux 2023 ou Amazon Linux 2 para a Amazon padrão. EMR AMI Para obter mais informações, consulte [Usando o Amazon Linux padrão AMI para a Amazon EMR](#).

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 709.1	4.14.348	23 de julho de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), AWS GovCloud China (Pequim), China AWS GovCloud (Ningxia), Ásia Pacífico (Hyderabad),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Oriente Médio (), Europa (EspanhaUAE), Europa (Zurique), Ásia-Pacífico (Melbourne), Israel (Tel Aviv), Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 223.0	4.14.336	8 de março de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia) , Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 131.0	4.14.336	14 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia) , Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 124.0	4.14.336	7 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia) , Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 109.0	4.14.334	24 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia), Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 218.0	4.14.330	2 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 206.0	4.14.330	22 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 116.0	4.14.328	11 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 101.0	4.14.327	17 de novembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelease Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 906.0	4.14.322	11 de setembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)

Versão 6.13.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 6.13.0 da Amazon. As alterações são referentes à versão 6.12.0. Para obter informações sobre o cronograma da versão, consulte o [Log de alterações 6.13.0](#).

Novos atributos

- 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.

Alterações, melhorias e problemas resolvidos

- A versão 6.13.0 aprimora o daemon de gerenciamento de EMR registros da Amazon para garantir que todos os registros sejam carregados regularmente para o Amazon S3 quando um comando de encerramento de cluster é emitido. Isso facilita o encerramento mais rápido do cluster.
- A versão 6.13.0 aprimora os recursos de gerenciamento de logs EMR da Amazon para garantir o upload consistente e oportuno de todos os arquivos de log para o Amazon S3. Isso beneficia especialmente os EMR clusters de longa duração.
- Quando você inicia um cluster com a versão de patch mais recente do Amazon EMR 5.36 ou superior, 6.6 ou superior ou 7.0 ou superior, a Amazon EMR usa a versão mais recente do Amazon Linux 2023 ou Amazon Linux 2 para a Amazon padrão. EMR AMI Para obter mais informações, consulte [Usando o Amazon Linux padrão AMI para a Amazon EMR](#).

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 709.1	4.14.348	23 de julho de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), AWS GovCloud China (Pequim), China AWS GovCloud (Ningxia), Ásia Pacífico (Hyderabad),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Oriente Médio (), Europa (EspanhaUAE), Europa (Zurique), Ásia-Pacífico (Melbourne), Israel (Tel Aviv), Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 223.0	4.14.336	8 de março de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia) , Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 131.0	4.14.336	14 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia) , Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 124.0	4.14.336	7 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia) , Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 109.0	4.14.334	24 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia) , Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 218.0	4.14.330	2 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 206.0	4.14.330	22 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 116.0	4.14.328	11 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 101.0	4.14.327	16 de novembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelease Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 020.1	4.14.326	7 de novembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 012.1	4.14.326	26 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 926.0	4.14.322	19 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 906.0	4.14.322	4 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 808.0	4.14.320	24 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)

Versão 6.12.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 6.12.0 da Amazon. As alterações são referentes à versão 6.11.0. Para obter informações sobre o cronograma da versão, consulte o [Log de alterações 6.12.0](#).

Novos atributos

- 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.
- A Amazon EMR lança 6.12.0 e versões superiores de suporte à LDAP integração com Apache Livy, Apache Hive through HiveServer 2 (HS2), Trino, Presto e Hue. Você também pode instalar o Apache Spark e o Apache Hadoop em um EMR cluster que usa 6.12.0 ou superior e configurá-los para uso. LDAP Para obter mais informações, consulte [Usar o Active Directory ou LDAP servidores para autenticação com a Amazon EMR](#).

Alterações, melhorias e problemas resolvidos

- As EMR versões 6.12.0 e superiores da Amazon oferecem suporte ao tempo de execução do Java 11 para o Flink. Para obter mais informações, consulte [Configurar o Flink para ser executado com o Java 11](#).
- A versão 6.12.0 adiciona um novo mecanismo de repetição ao fluxo de trabalho de escalabilidade de EMR clusters para clusters que executam o Presto ou o Trino. Essa melhoria reduz o risco de que o redimensionamento do cluster fique paralisado indefinidamente devido a uma única falha na operação de redimensionamento. Ela também aprimora a utilização dos clusters, porque seu cluster aumenta e reduz a escala verticalmente com mais rapidez.

- A versão 6.12.0 corrige um problema em que as operações de redução da escala verticalmente do cluster podem ficar paralisadas quando um nó central que está passando por uma desativação tranquila se torna não íntegro por qualquer motivo antes de ser totalmente desativado.
- A versão 6.12.0 melhora a lógica de redução do cluster para que seu cluster não tente reduzir os nós principais abaixo da configuração do fator de HDFS replicação do cluster. Isso se alinha aos seus requisitos de redundância de dados e reduz a probabilidade de uma operação de escalabilidade paralisar.
- A versão 6.12.0 aprimora o desempenho e a eficiência do serviço de monitoramento de saúde da Amazon, EMR aumentando a velocidade com que ele registra as mudanças de estado das instâncias. Essa melhoria reduz a probabilidade de degradação do desempenho dos nós do cluster que estão executando várias ferramentas de cliente ou aplicações de terceiros personalizadas.
- A versão 6.12.0 melhora o desempenho do daemon de gerenciamento de logs no cluster para a Amazon. EMR Como resultado, há menos chances de degradação do desempenho com EMR clusters que executam etapas com alta simultaneidade.
- Com a EMR versão 6.12.0 da Amazon, o daemon de gerenciamento de registros foi atualizado para identificar todos os registros que estão em uso ativo com identificadores de arquivos abertos no armazenamento da instância local e os processos associados. Essa atualização garante que a Amazon exclua EMR adequadamente os arquivos e recupere o espaço de armazenamento depois que os registros forem arquivados no Amazon S3.
- A versão 6.12.0 inclui um aprimoramento do daemon de gerenciamento de logs que exclui diretórios de etapas vazios e não utilizados no sistema de arquivos de cluster local. Um número excessivamente grande de diretórios vazios pode degradar o desempenho dos EMR daemons da Amazon e resultar na sobreutilização do disco.
- A versão 6.12.0 permite a rotação de registros para registros YARN do Timeline Server. Isso minimiza os cenários de utilização excessiva do disco, especialmente para clusters de execução prolongada.
- O tamanho padrão do volume raiz aumentou para 15 GB no Amazon EMR 6.10.0 e superior. O tamanho padrão do volume raiz das versões anteriores é de 10 GB.
- Quando você inicia um cluster com a versão de patch mais recente do Amazon EMR 5.36 ou superior, 6.6 ou superior ou 7.0 ou superior, a Amazon EMR usa a versão mais recente do Amazon Linux 2023 ou Amazon Linux 2 para a Amazon padrão. EMR AMI Para obter mais informações, consulte [Usando o Amazon Linux padrão AMI para a Amazon EMR](#).

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 709.1	4.14.348	23 de julho de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), AWS GovCloud China (Pequim), China AWS GovCloud (Ningxia), Ásia Pacífico (Hyderabad),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Oriente Médio (), Europa (EspanhaUAE), Europa (Zurique), Ásia-Pacífico (Melbourne), Israel (Tel Aviv), Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 223.0	4.14.336	8 de março de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia) , Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 131.0	4.14.336	14 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia) , Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 124.0	4.14.336	7 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia), Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 109.0	4.14.334	24 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelease Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia) , Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 218.0	4.14.330	2 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 206.0	4.14.330	22 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelease Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 116.0	4.14.328	11 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 101.0	4.14.327	16 de novembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelease Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 020.1	4.14.326	7 de novembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 012.1	4.14.326	26 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 926.0	4.14.322	19 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 906.0	4.14.322	4 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 822.0	4.14.322	30 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 808.0	4.14.320	24 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 727.0	4.14.320	14 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 719.0	4.14.320	2 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 628.0	4.14.318	12 de julho de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Oriente Médio (), Canadá (Central) UAE

Versão 6.11.1

As notas de lançamento a seguir incluem informações sobre a EMR versão 6.11.1 da Amazon. As alterações são referentes à versão 6.11.0. Para obter informações sobre o cronograma da versão, consulte o [Log de alterações 6.11.1](#).

Alterações, melhorias e problemas resolvidos

- Devido à contenção de bloqueio, um nó pode entrar em um deadlock se for adicionado ou removido ao mesmo tempo em que tenta ser desativado. Como resultado, o Hadoop Resource Manager (YARN) deixa de responder e afeta todos os contêineres recebidos e em execução no momento.
- Esta versão inclui uma alteração que permite que clusters de alta disponibilidade se recuperem de um estado de falha após a reinicialização.
- Esta versão inclui correções de segurança para Hue e HBase
- Esta versão corrige um problema em que clusters que estão executando cargas de trabalho no Spark com a Amazon EMR podem receber silenciosamente resultados incorretos com `contains`, `startsWith`, e `endsWith` `like`. Esse problema ocorre quando você usa as expressões em campos particionados que têm metadados no Amazon EMR Hive3 Metastore Server (). HMS
- Esta versão corrige um problema de limitação no lado do Glue quando não há funções definidas pelo usuário (). UDF
- Esta versão corrige um problema que exclui registros de contêineres pelo serviço de agregação de registros de nós antes que o carregador de registros possa enviá-los para o S3 em caso de descomissionamento. YARN
- Esta versão corrige um problema com as métricas do FairShare Scheduler quando o Node Label está habilitado para o Hadoop.
- Esta versão corrige um problema que afetou a performance do Spark quando você definiu um valor de `true` padrão para a configuração `spark.yarn.heterogeneousExecutors.enabled` no `spark-defaults.conf`.
- Esta versão corrige um problema com a falha do Reduce Task em ler dados embaralhados. O problema causou falhas na consulta do Hive com um erro de memória corrompida.
- Esta versão adiciona um novo mecanismo de repetição ao fluxo de trabalho de escalabilidade de EMR clusters para clusters que executam o Presto ou o Trino. Essa melhoria reduz o risco de que o redimensionamento do cluster fique paralisado indefinidamente devido a uma única falha

na operação de redimensionamento. Ela também aprimora a utilização dos clusters, porque seu cluster aumenta e reduz a escala verticalmente com mais rapidez.

- Esta versão melhora a lógica de redução do cluster para que seu cluster não tente reduzir os nós principais abaixo da configuração do fator de HDFS replicação do cluster. Isso se alinha aos seus requisitos de redundância de dados e reduz a probabilidade de uma operação de escalabilidade paralisar.
- O daemon de gerenciamento de logs foi atualizado para identificar todos os logs que estão em uso ativo com identificadores de arquivos abertos no armazenamento da instância local e nos processos associados. Essa atualização garante que a Amazon exclua EMR adequadamente os arquivos e recupere o espaço de armazenamento depois que os registros forem arquivados no Amazon S3.
- Esta versão inclui um aprimoramento do daemon de gerenciamento de logs que exclui diretórios de etapas vazios e não utilizados no sistema de arquivos de cluster local. Um número excessivamente grande de diretórios vazios pode degradar o desempenho dos EMR daemons da Amazon e resultar na sobreutilização do disco.
- Quando você inicia um cluster com a versão de patch mais recente do Amazon EMR 5.36 ou superior, 6.6 ou superior ou 7.0 ou superior, a Amazon EMR usa a versão mais recente do Amazon Linux 2023 ou Amazon Linux 2 para a Amazon padrão. EMR AMI Para obter mais informações, consulte [Usando o Amazon Linux padrão AMI para a Amazon EMR](#).

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 709.1	4.14.348	23 de julho de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			<p>Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), AWS GovCloud China (Pequim), China AWS GovCloud (Ningxia), Ásia Pacífico (Hyderabad), Oriente Médio (), Europa (EspanhaUAE), Europa (Zurique), Ásia-Pacífico (Melbourne), Israel (Tel Aviv), Oeste do Canadá (Calgary)</p>

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 223.0	4.14.336	8 de março de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelease Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia) , Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 131.0	4.14.336	14 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelease Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia) , Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 124.0	4.14.336	7 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia) , Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 109.0	4.14.334	24 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelease Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China (Ningxia), Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 218.0	4.14.330	2 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 206.0	4.14.330	22 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 116.0	4.14.328	11 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 101.0	4.14.327	16 de novembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelease Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 020.1	4.14.326	7 de novembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 012.1	4.14.326	26 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 926.0	4.14.322	19 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Canadá (UAE Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Israel (Tel Aviv), AWS GovCloud (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 906.0	4.14.322	4 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 822.0	4.14.322	30 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 808.0	4.14.320	24 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)
2.0.2023 727.0	4.14.320	14 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central)

Versão 6.11.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 6.11.0 da Amazon. As alterações são referentes à versão 6.10.0. Para obter informações sobre o cronograma da versão, consulte o [log de alterações](#).

Novos atributos

- O Amazon EMR 6.11.0 é compatível com 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 1.2.0-amzn-0, Trino 410-amzn-0 e PrestoDB 0.279-amzn-0. CUDA

Alterações, melhorias e problemas resolvidos

- Com o Amazon EMR 6.11.0, o conector do DynamoDB foi atualizado para a versão 5.0.0. A versão 5.0.0 usa AWS SDK for Java 2.x. As versões anteriores usavam AWS SDK for Java 1.x. Como resultado dessa atualização, recomendamos que você teste seu código antes de usar o conector do DynamoDB com o Amazon 6.11. EMR
- Quando o conector do DynamoDB para EMR Amazon 6.11.0 chama o serviço do DynamoDB, ele usa o valor da região que você fornece para a propriedade `dynamodb.endpoint`. Recomendamos que você também configure `dynamodb.region` quando usar `dynamodb.endpoint` e que ambas as propriedades tenham como destino a mesma Região da AWS. Se você usar `dynamodb.endpoint` e não configurar `dynamodb.region`, o conector do DynamoDB para EMR Amazon 6.11.0 retornará uma exceção de região inválida e tentará reconciliar suas Região da AWS informações do serviço de metadados de instância da Amazon (`EC2 IMDS`). Se o conector não conseguir recuperar a região `IMDS`, o padrão é Leste dos EUA (Norte da Virgínia) (`us-east-1`). O erro a seguir é um exemplo da exceção de região inválida que você pode obter se não configurar adequadamente a `dynamodb.region` propriedade:
`error software.amazon.awssdk.services.dynamodb.model.DynamoDbException: Credential should be scoped to a valid region.` Para obter mais informações sobre as classes afetadas pela AWS SDK for Java atualização para 2.x, consulte o commit [Upgrade AWS SDK for Java from 1.x to 2.x \(#175\)](#) no GitHub repositório do conector Amazon EMR - DynamoDB.
- Esta versão corrige um problema em que os dados da coluna se tornam NULL quando você usa o Delta Lake para armazenar dados da tabela Delta no Amazon S3 após a operação de renomeação da coluna. Para obter mais informações sobre esse atributo experimental no Delta Lake, consulte [Operação de renomeação de coluna](#) no Guia do usuário do Delta Lake.


- A versão 6.11.0 corrige um problema que pode ocorrer quando você cria um nó de borda ao replicar um dos nós primários de um cluster com vários nós primários. O nó de borda replicado pode causar atrasos nas operações de redução da escala verticalmente ou resultar em alta utilização de memória nos nós primários. Para obter mais informações sobre como criar um nó de borda para se comunicar com seu EMR cluster, consulte [Edge Node Creator](#) no `aws-samples` repositório em GitHub.
- A versão 6.11.0 melhora o processo de automação que a Amazon EMR usa para remontar volumes da EBS Amazon em uma instância após uma reinicialização.
- A versão 6.11.0 corrige um problema que resultou em lacunas intermitentes nas métricas do Hadoop que a Amazon publica na Amazon. EMR CloudWatch
- A versão 6.11.0 corrige um problema com EMR clusters em que uma atualização no arquivo de YARN configuração que contém a lista de exclusão de nós do cluster é interrompida devido à utilização excessiva do disco. A atualização incompleta impede futuras operações de redução da escala verticalmente do cluster. Esta versão garante que o cluster permaneça íntegro e que as operações de escalabilidade funcionem conforme esperado.
- O tamanho padrão do volume raiz aumentou para 15 GB no Amazon EMR 6.10.0 e superior. O tamanho padrão do volume raiz das versões anteriores é de 10 GB.
- O Hadoop 3.3.3 introduziu uma alteração em YARN ([YARN-9608](#)) que mantém os nós em que os contêineres eram executados em um estado de descomissionamento até que o aplicativo seja concluído. Essa alteração garante que dados locais, como dados embaralhados, não sejam perdidos e que você não precise executar o trabalho novamente. Essa abordagem também pode levar à subutilização de recursos em clusters com ou sem o ajuste de escala gerenciado habilitado.

Com as EMR versões 6.11.0 e superiores da Amazon, bem como 6.8.1, 6.9.1 e 6.10.1, o valor de `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` é definido como `in` para resolver esse problema. `false` `yarn-site.xml`

Embora a correção resolva os problemas introduzidos pelo YARN -9608, ela pode fazer com que as tarefas do Hive falhem devido à perda aleatória de dados em clusters com escalabilidade gerenciada habilitada. Reduzimos esse risco nesta versão também ao configurar `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-shuffle-data` para workloads do Hive. Essa configuração só está disponível nas EMR versões 6.11.0 e superiores da Amazon.

- Quando você inicia um cluster com a versão de patch mais recente do Amazon EMR 5.36 ou superior, 6.6 ou superior ou 7.0 ou superior, a Amazon EMR usa a versão mais recente do

Amazon Linux 2023 ou Amazon Linux 2 para a Amazon padrão. EMR AMI Para obter mais informações, consulte [Usando o Amazon Linux padrão AMI para a Amazon EMR](#).

 Note

Esta versão não recebe mais AMI atualizações automáticas, pois foi sucedida por mais 1 lançamento de patch. A versão de patch é indicada pelo número após o segundo ponto decimal (6.8.**1**). Para ver se você está usando a versão de patch mais recente, verifique as versões disponíveis no [Guia de lançamento](#) ou verifique o menu suspenso de EMRlançamentos da Amazon ao criar um cluster no console ou usar a ação [ListReleaseLabels](#)API ou [list-release-labels](#)CLI. Para receber atualizações sobre novos lançamentos, assine o RSS feed do [What's new?](#) página.

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 709.1	4.14.348	23 de julho de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), AWS GovCloud China (Pequim), China AWS GovCloud (Ningxia), Ásia Pacífico (Hyderabad),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Oriente Médio (), Europa (EspanhaUAE), Europa (Zurique), Ásia-Pacífico (Melbourne), Israel (Tel Aviv), Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 808.0	4.14.320	24 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), Israel (Tel Aviv).

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 727.0	4.14.320	14 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 719.0	4.14.320	2 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)
2.0.2023 628.0	4.14.318	12 de julho de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Zurique), Europa (Milão), Europa (Espanha), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad)), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio () UAE

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 612.0	4.14.314	23 de junho de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Zurique), Europa (Milão), Europa (Espanha), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad)), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio () UAE

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 504.1	4.14.313	16 de maio de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Oriente Médio (), Canadá (Central) UAE

Versão 6.10.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 6.10.0 da Amazon. As alterações são referentes à versão 6.9.0. Para obter informações sobre o cronograma da versão, consulte o [log de alterações](#).

Novos atributos

- O Amazon EMR 6.10.0 é compatível com 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 e PrestoDB 0.278.1.
- O Amazon EMR 6.10.0 inclui um conector Trino-Hudi nativo que fornece acesso de leitura aos dados nas tabelas Hudi. Você pode ativar o conector com `trino-cli --catalog hudi` e configurar o conector de acordo com suas necessidades com `trino-connector-hudi`. A integração nativa com a Amazon EMR significa que você não precisa mais usar `trino-connector-hive` para consultar tabelas Hudi. Para obter uma lista das configurações compatíveis com o novo conector, consulte a página do [conector do Hudi](#) na documentação do Trino.
- As EMR versões 6.10.0 e superiores da Amazon oferecem suporte à integração do Apache Zeppelin com o Apache Flink. Consulte [Usar trabalhos do Flink pelo Zeppelin no Amazon EMR](#) Para mais informações.

Problemas conhecidos

- O Hadoop 3.3.3 introduziu uma alteração em YARN ([YARN-9608](#)) que mantém os nós em que os contêineres eram executados em um estado de descomissionamento até que o aplicativo seja concluído. Essa alteração garante que dados locais, como dados embaralhados, não sejam perdidos e que você não precise executar o trabalho novamente. Essa abordagem também pode levar à subutilização de recursos em clusters com ou sem o ajuste de escala gerenciado habilitado.

Para contornar esse problema no Amazon EMR 6.10.0, você pode definir o valor de `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications to false` in `yarn-site.xml` Nas EMR versões 6.11.0 e superiores da Amazon, bem como nas 6.8.1, 6.9.1 e 6.10.1, a configuração é definida como padrão para resolver esse problema. `false`

Alterações, melhorias e problemas resolvidos

- O Amazon EMR 6.10.0 remove a dependência `minimal-json.jar` da [integração do Amazon Redshift com o Apache Spark e adiciona automaticamente os jars relacionados ao Spark-Redshift](#) necessários ao caminho da classe executora do Spark:, e. `spark-redshift.jar` `spark-avro.jar` `RedshiftJDBC.jar`
- A versão 6.10.0 aprimora o daemon de gerenciamento de registros no cluster para monitorar pastas de log adicionais em seu cluster. EMR Essa melhoria minimiza os cenários de utilização excessiva do disco.
- A versão 6.10.0 reinicia automaticamente o daemon de gerenciamento de logs no cluster quando ele é interrompido. Essa melhoria reduz o risco de os nós parecerem não íntegros devido à utilização excessiva do disco.
- O Amazon EMR 6.10.0 oferece suporte a endpoints regionais para EMRFS mapeamento de usuários.
- O tamanho padrão do volume raiz aumentou para 15 GB no Amazon EMR 6.10.0 e superior. O tamanho padrão do volume raiz das versões anteriores é de 10 GB.
- A versão 6.10.0 corrige um problema que fazia com que as tarefas do Spark parassem quando todos os executores restantes do Spark estavam em um host de desativação com o gerente de recursos. YARN
- Com o Amazon EMR 6.6.0 a 6.9.x, INSERT as consultas com partição dinâmica e uma cláusula ORDER BY ou SORT BY sempre terão dois redutores. Esse problema é causado pela OSS alteração [HIVE-20703](#), que coloca a otimização dinâmica da partição de classificação sob decisão baseada em custos. Se sua workload não exigir a classificação de partições dinâmicas, recomendamos que você defina a propriedade `hive.optimize.sort.dynamic.partition.threshold` como `-1` para desabilitar o novo atributo e obter o número de redutores calculado corretamente. Esse problema foi corrigido no OSS Hive como parte do [HIVE-22269](#) e foi corrigido no Amazon 6.10.0. EMR
- Quando você inicia um cluster com a versão de patch mais recente do Amazon EMR 5.36 ou superior, 6.6 ou superior ou 7.0 ou superior, a Amazon EMR usa a versão mais recente do Amazon Linux 2023 ou Amazon Linux 2 para a Amazon padrão. EMR AMI Para obter mais informações, consulte [Usando o Amazon Linux padrão AMI para a Amazon EMR](#).

Note

Esta versão não recebe mais AMI atualizações automáticas, pois foi sucedida por mais 1 lançamento de patch. A versão de patch é indicada pelo número após o segundo

ponto decimal (6.8.1). Para ver se você está usando a versão de patch mais recente, verifique as versões disponíveis no [Guia de lançamento](#) ou verifique o menu suspenso de EMRlançamentos da Amazon ao criar um cluster no console ou usar a ação [ListReleaseLabels](#)API ou [list-release-labels](#)CLI. Para receber atualizações sobre novos lançamentos, assine o RSS feed do [What's new?](#) página.

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 709.1	4.14.348	23 de julho de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), AWS GovCloud China (Pequim), China AWS GovCloud (Ningxia), Ásia Pacífico (Hyderabad),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Oriente Médio (), Europa (EspanhaUAE), Europa (Zurique), Ásia-Pacífico (Melbourne), Israel (Tel Aviv), Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 808.0	4.14.320	24 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), Israel (Tel Aviv).

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 727.0	4.14.320	14 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 719.0	4.14.320	2 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)
2.0.2023 628.0	4.14.318	12 de julho de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Zurique), Europa (Milão), Europa (Espanha), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad)), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio () UAE

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 612.0	4.14.314	23 de junho de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Zurique), Europa (Milão), Europa (Espanha), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad)), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio () UAE

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 504.1	4.14.313	16 de maio de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Zurique), Europa (Milão), Europa (Espanha), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad)), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio () UAE

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 418.0	4.14.311	3 de maio de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Zurique), Europa (Milão), Europa (Espanha), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad)), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio () UAE

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 404.1	4.14.311	18 de abril de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul)), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio () UAE
2.0.2023 404.0	4.14.311	10 de abril de 2023	Leste dos EUA (Norte da Virgínia), Europa (Paris)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 320.0	4.14.309	30 de março de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul)), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio () UAE

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 207.0	4.14.304	22 de fevereiro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul)), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio () UAE

Versão 6.9.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 6.9.0 da Amazon. As alterações são relativas à EMR versão 6.8.0 da Amazon. Para obter informações sobre o cronograma da versão, consulte o [log de alterações](#).

Novos atributos

- A EMR versão 6.9.0 da Amazon é compatível com Apache Spark RAPIDS 22.08.0, Apache Hudi 0.12.1, Apache Iceberg 0.14.1, Trino 398 e Tez 0.10.2.
- A EMR versão 6.9.0 da Amazon inclui um novo aplicativo de código aberto, 2.1.0. [Delta Lake](#)
- A integração do Amazon Redshift para o Apache Spark está incluída nas EMR versões 6.9.0 e posteriores da Amazon. Anteriormente uma ferramenta de código aberto, a integração nativa é um conector do Spark que você pode usar para criar aplicações do Apache Spark que realizam a leitura e a gravação de dados no Amazon Redshift e no Amazon Redshift sem servidor. Para obter mais informações, consulte [Usando a integração do Amazon Redshift para o Apache Spark com a Amazon EMR](#).
- A EMR versão 6.9.0 da Amazon adiciona suporte ao arquivamento de registros no Amazon S3 durante a redução da escala do cluster. Anteriormente, só era possível arquivar arquivos de log no Amazon S3 durante o encerramento do cluster. A nova capacidade garante que os arquivos de log gerados no cluster persistam no Amazon S3 mesmo após o encerramento do nó. Para obter mais informações, consulte [Configurar registro em log e depuração do cluster](#).
- Para dar suporte a consultas de longa execução, o Trino agora inclui um mecanismo de execução tolerante a falhas. A execução tolerante a falhas atenua as falhas nas consultas ao tentar novamente as consultas com falha ou as tarefas dos seus componentes. Para obter mais informações, consulte [Execução tolerante a falhas no Trino](#).
- Você pode usar o Apache Flink na Amazon EMR para unificar BATCH e STREAM processar tabelas do Apache Hive ou metadados de qualquer fonte de tabela do Flink, como Iceberg, Kinesis ou Kafka. Você pode especificar o AWS Glue Data Catalog como metastore para o Flink usando o AWS Management Console,, AWS CLI ou Amazon. EMR API Para obter mais informações, consulte [Configurando o Flink na Amazon EMR](#).
- Agora você pode especificar AWS Identity and Access Management (IAM) funções de tempo de execução e controle de acesso AWS Lake Formation baseado para consultas do Apache Spark, Apache Hive e Presto na Amazon EMR em EC2 clusters com o Amazon Studio. SageMaker Para obter mais informações, consulte [Configurar funções de tempo de execução para EMR as etapas da Amazon](#).

Problemas conhecidos

- Para a EMR versão 6.9.0 da Amazon, o Trino não funciona em clusters habilitados para o Apache Ranger. Se você precisar usar o Trino com o Ranger, entre em contato com o [AWS Support](#).

- Se você usar a integração do Amazon Redshift para Apache Spark e tiver um time, timetz, timestamp ou timestamptz com precisão de microssegundos no formato Parquet, o conector arredondará os valores de tempo para o valor de milissegundo mais próximo. Como solução alternativa, use o parâmetro `unload_s3_format` do formato de descarregamento de texto.
- Quando você usa o Spark com a formatação de localização de partições do Hive para ler dados no Amazon S3 e executa o Spark nas EMR versões 5.30.0 a 5.36.0 e 6.2.0 a 6.9.0 da Amazon, você pode encontrar um problema que impede seu cluster de ler os dados corretamente. Isso poderá acontecer se suas partições tiverem todas as características a seguir:
 - Duas ou mais partições são verificadas na mesma tabela.
 - Pelo menos um caminho de diretório de partição é um prefixo de pelo menos outro caminho de diretório de partição, por exemplo, `s3://bucket/table/p=a` é um prefixo de `s3://bucket/table/p=a b`.
 - O primeiro caractere que segue o prefixo no outro diretório de partição tem um valor UTF -8 menor que o / caractere (U+002F). Por exemplo, o caractere de espaço (U+0020) que ocorre entre a e b em `s3://bucket/table/p=a b` se enquadra nessa categoria. Observe que existem 14 outros caracteres que não são de controle: !"#\$%&'()*+,-. Para obter mais informações, consulte a [tabela de codificação UTF -8 e caracteres Unicode](#).

Como solução alternativa para esse problema, defina a configuração `spark.sql.sources.fastS3PartitionDiscovery.enabled` como `false` na classificação `spark-defaults`.

- As conexões com EMR clusters da Amazon a partir do Amazon SageMaker Studio podem falhar intermitentemente com um código de resposta 403 Forbidden. Esse erro ocorre quando a configuração da IAM função no cluster leva mais de 60 segundos. Como solução alternativa, você pode instalar um EMR patch da Amazon para permitir novas tentativas e aumentar o tempo limite para um mínimo de 300 segundos. Use as etapas a seguir para aplicar a ação de bootstrap quando iniciar o cluster.

1. Baixe o script e os RPM arquivos de bootstrap do Amazon URIs S3 a seguir.

```
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. Carregue os arquivos da etapa anterior em um bucket do Amazon S3 de sua propriedade. O bucket deve estar no mesmo Região da AWS local em que você planeja iniciar o cluster.

3. Inclua a seguinte ação de bootstrap ao iniciar seu EMR cluster. Substituir *bootstrap_URI* e *RPM_URI* com o correspondente URIs do Amazon S3.

```
--bootstrap-actions "Path=bootstrap_URI,Args=[RPM_URI]"
```

- Com as EMR versões 5.36.0 e 6.6.0 a 6.9.0 da Amazon, SecretAgent os componentes do RecordServer serviço podem sofrer perda de dados de log devido a uma configuração incorreta do padrão de nome de arquivo nas propriedades do Log4j2. A configuração incorreta faz com que os componentes gerem somente um arquivo de log por dia. Quando a estratégia de rotação ocorre, ela substitui o arquivo existente em vez de gerar um novo arquivo de log, conforme esperado. Como solução alternativa, use uma ação de bootstrap para gerar arquivos de log a cada hora e acrescentar um número inteiro de incremento automático no nome do arquivo para lidar com a rotação.

Para as versões EMR 6.6.0 a 6.9.0 da Amazon, use a seguinte ação de bootstrap ao iniciar um cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Para o Amazon EMR 5.36.0, use a seguinte ação de bootstrap ao iniciar um cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

- O Apache Flink fornece FileSystem conectores nativos S3 FileSystem e Hadoop, que permitem que os aplicativos criem FileSink e gravem os dados no Amazon S3. Isso FileSink falha com uma das duas exceções a seguir.

```
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:?]

```

Como solução alternativa, você pode instalar um EMR patch da Amazon, que corrige o problema acima no Flink. Para aplicar a ação de bootstrap quando iniciar o cluster, execute as etapas a seguir.

1. Baixe o flink-rpm no bucket Amazon S3. Seu RPM caminho é `s3://DOC-EXAMPLE-BUCKET/rpms/flink/`.
2. Faça o download do script e dos RPM arquivos de bootstrap do Amazon S3 usando o seguinte. URI `regionName` Substitua pelo Região da AWS local em que você planeja iniciar o cluster.

```

s3://emr-data-access-control-regionName/customer-bootstrap-actions/gcsc/replace-
rpms.sh

```

3. O Hadoop 3.3.3 introduziu uma alteração em YARN ([YARN-9608](#)) que mantém os nós em que os contêineres eram executados em um estado de descomissionamento até que o aplicativo seja concluído. Essa alteração garante que dados locais, como dados embaralhados, não sejam perdidos e que você não precise executar o trabalho novamente. Na Amazon EMR 6.8.0 e 6.9.0, essa abordagem também pode levar à subutilização de recursos em clusters com ou sem escalabilidade gerenciada habilitada.

Com o [Amazon EMR 6.10.0](#), há uma solução alternativa para esse problema: definir o valor de `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications to in. false` `yarn-site.xml` Nas EMR versões 6.11.0 e superiores da Amazon, bem como nas 6.8.1, 6.9.1 e 6.10.1, a configuração é definida como padrão para resolver esse problema. `false`

Alterações, melhorias e problemas resolvidos

- Para a Amazon EMR versão 6.9.0 e posterior, todos os componentes instalados pela Amazon EMR que usam bibliotecas Log4j usam Log4j versão 2.17.1 ou posterior.
- Quando você usa o conector do DynamoDB com o Spark nas versões 6.6.0, 6.7.0 e 6.8.0 da EMR Amazon, todas as leituras da sua tabela retornam um resultado vazio, mesmo que a divisão de entrada faça referência a dados não vazios. A EMR versão 6.9.0 da Amazon corrige esse problema.

- O Amazon EMR 6.9.0 adiciona suporte limitado ao controle de acesso baseado em Lake Formation com o Apache Hudi ao ler dados usando o Spark. SQL O suporte é para SELECT consultas usando o Spark SQL e é limitado ao controle de acesso em nível de coluna. Para obter mais informações, consulte [Hudi e Lake Formation](#).
- Quando você usa o Amazon EMR 6.9.0 para criar um cluster Hadoop com [Node Labels](#) habilitados, as [YARNmétricas API](#) retornam informações agregadas em todas as partições, em vez da partição padrão. Para obter mais informações, consulte [YARN-11414](#).
- Com a EMR versão 6.9.0 da Amazon, atualizamos o Trino para a versão 398, que usa Java 17. A versão anterior compatível do Trino para Amazon EMR 6.8.0 era a Trino 388 em execução no Java 11. Para obter mais informações sobre essa alteração, consulte [Atualizações do Trino para Java 17](#) no blog do Trino.
- Esta versão corrige um problema de incompatibilidade de sequência de tempo entre o Apache BigTop e a Amazon EMR na sequência de inicialização EC2 do cluster. Essa incompatibilidade de sequência de tempo ocorre quando um sistema tenta realizar duas ou mais operações ao mesmo tempo em vez de fazê-las na sequência correta. Como resultado, determinadas configurações de cluster apresentaram tempos limite de inicialização da instância e tempos de inicialização do cluster mais lentos.
- Quando você inicia um cluster com a versão de patch mais recente do Amazon EMR 5.36 ou superior, 6.6 ou superior ou 7.0 ou superior, a Amazon EMR usa a versão mais recente do Amazon Linux 2023 ou Amazon Linux 2 para a Amazon padrão. EMR AMI Para obter mais informações, consulte [Usando o Amazon Linux padrão AMI para a Amazon EMR](#).

Note

Esta versão não recebe mais AMI atualizações automáticas, pois foi sucedida por mais 1 lançamento de patch. A versão de patch é indicada pelo número após o segundo ponto decimal (6.8.**1**). Para ver se você está usando a versão de patch mais recente, verifique as versões disponíveis no [Guia de lançamento](#) ou verifique o menu suspenso de EMRlançamentos da Amazon ao criar um cluster no console ou usar a ação [ListReleaseLabelsAPI](#) ou [list-release-labelsCLI](#). Para receber atualizações sobre novos lançamentos, assine o RSS feed do [What's new?](#) página.

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 709.1	4.14.348	23 de julho de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), AWS GovCloud China (Pequim), China AWS GovCloud (Ningxia), Ásia Pacífico (Hyderabad),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			Oriente Médio (), Europa (EspanhaUAE), Europa (Zurique), Ásia-Pacífico (Melbourne), Israel (Tel Aviv), Oeste do Canadá (Calgary)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 808.0	4.14.320	24 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), Israel (Tel Aviv).

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 727.0	4.14.320	14 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsReleaseLabel (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 719.0	4.14.320	2 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			(), Canadá (UAE Central), Israel (Tel Aviv)
2.0.2023 628.0	4.14.318	12 de julho de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 612.0	4.14.314	23 de junho de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 504.1	4.14.313	16 de maio de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 418.0	4.14.311	3 de maio de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 404.1	4.14.311	18 de abril de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)
2.0.2023 404.0	4.14.311	10 de abril de 2023	Leste dos EUA (Norte da Virgínia), Europa (Paris)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 320.0	4.14.309	30 de março de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 307.0	4.14.305	15 de março de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 207.0	4.14.304	22 de fevereiro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 210.1	4.14.301	12 de janeiro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 103.3	4.14.296	5 de dezembro de 2022	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

Versão 6.8.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 6.8.0 da Amazon. As alterações são referentes à versão 6.7.0.

Novos atributos

- O recurso Amazon EMR Steps agora é compatível com endpoint e JDBC /clients Apache Livy. ODBC Para obter mais informações, consulte [Configurar funções de tempo de execução para EMR as etapas da Amazon](#).
- A EMR versão 6.8.0 da Amazon vem com a versão 2.4.12 do ApacheHBase. Com essa HBase versão, você pode arquivar e excluir suas HBase tabelas. O processamento de arquivos do Amazon S3 renomeia todos os arquivos da tabela para o diretório de arquivos. Isso pode ser um processo custoso e demorado. Agora, você pode pular o processamento de arquivos e rapidamente eliminar e excluir tabelas grandes. Para obter mais informações, consulte [Usando a HBase concha](#).

Problemas conhecidos

- O Hadoop 3.3.3 introduziu uma alteração em YARN ([YARN-9608](#)) que mantém os nós em que os contêineres eram executados em um estado de descomissionamento até que o aplicativo seja concluído. Essa alteração garante que dados locais, como dados embaralhados, não sejam perdidos e que você não precise executar o trabalho novamente. Na Amazon EMR 6.8.0 e 6.9.0, essa abordagem também pode levar à subutilização de recursos em clusters com ou sem escalabilidade gerenciada habilitada.

Com o [Amazon EMR 6.10.0](#), há uma solução alternativa para esse problema: definir o valor de `yarn.resourcemanager.decommissioning-nodes-watcher.wait-for-applications` to `in. false` `yarn-site.xml` Nas EMR versões 6.11.0 e superiores da Amazon, bem como nas 6.8.1, 6.9.1 e 6.10.1, a configuração é definida como padrão para resolver esse problema. `false`

Alterações, melhorias e problemas resolvidos

- Quando a Amazon EMR lança 6.5.0, 6.6.0 ou 6.7.0 leu as tabelas do Apache Phoenix por meio do shell do Apache Spark, a Amazon produziu uma. EMR `NoSuchMethodError` A EMR versão 6.8.0 da Amazon corrige esse problema.
- A EMR versão 6.8.0 da Amazon vem com o [Apache Hudi](#) 0.11.1; no entanto, os clusters da EMR Amazon 6.8.0 também são compatíveis com o código aberto do Hudi 0.12.0. `hudi-spark3.3-bundle_2.12`
- A EMR versão 6.8.0 da Amazon vem com o Apache Spark 3.3.0. Esta versão do Spark usa o Apache Log4j 2 e o arquivo `log4j2.properties` para configurar o Log4j nos processos do

Spark. Se você usa o Spark no cluster ou cria EMR clusters com parâmetros de configuração personalizados e deseja atualizar para a EMR versão 6.8.0 da Amazon, você deve migrar para a nova classificação de `spark-log4j2` configuração e formato de chave do Apache Log4j 2. Para obter mais informações, consulte [Migrar do Apache Log4j 1.x para Log4j 2.x](#).

- Quando você inicia um cluster com a versão de patch mais recente do Amazon EMR 5.36 ou superior, 6.6 ou superior ou 7.0 ou superior, a Amazon EMR usa a versão mais recente do Amazon Linux 2023 ou Amazon Linux 2 para a Amazon padrão. EMR AMI Para obter mais informações, consulte [Usando o Amazon Linux padrão AMI para a Amazon EMR](#).

Note

Esta versão não recebe mais AMI atualizações automáticas, pois foi sucedida por mais 1 lançamento de patch. A versão de patch é indicada pelo número após o segundo ponto decimal (6.8.**1**). Para ver se você está usando a versão de patch mais recente, verifique as versões disponíveis no [Guia de lançamento](#) ou verifique o menu suspenso de EMRlançamentos da Amazon ao criar um cluster no console ou usar a ação [ListReleaseLabels](#)API ou [list-release-labels](#)CLI. Para receber atualizações sobre novos lançamentos, assine o RSS feed do [What's new?](#) página.

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 709.1	4.14.348	23 de julho de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			<p>Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), AWS GovCloud China (Pequim), China AWS GovCloud (Ningxia), Ásia Pacífico (Hyderabad), Oriente Médio (), Europa (EspanhaUAE), Europa (Zurique), Ásia-Pacífico (Melbourne), Israel (Tel Aviv)</p>

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 808.0	4.14.320	24 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 727.0	4.14.320	14 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central),

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 719.0	4.14.320	2 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Melbourne), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (UAE Central), Canadá (UAE Central)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 628.0	4.14.318	12 de julho de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 612.0	4.14.314	23 de junho de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 504.1	4.14.313	16 de maio de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 418.0	4.14.311	3 de maio de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 404.1	4.14.311	18 de abril de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)
2.0.2023 404.0	4.14.311	10 de abril de 2023	Leste dos EUA (Norte da Virgínia), Europa (Paris)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 320.0	4.14.309	30 de março de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 307.0	4.14.305	15 de março de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 207.0	4.14.304	22 de fevereiro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 119.1	4.14.301	3 de fevereiro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 210.1	4.14.301	22 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 103.3	4.14.296	5 de dezembro de 2022	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 004.0	4.14.294	2 de novembro de 2022	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 912.1	4.14.291	6 de setembro de 2022	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

Problemas conhecidos

- Quando você usa o conector do DynamoDB com o Spark nas versões 6.6.0, 6.7.0 e 6.8.0 da EMR Amazon, todas as leituras da sua tabela retornam um resultado vazio, mesmo que a divisão de entrada faça referência a dados não vazios. Isso ocorre porque o Spark 3.2.0 define `spark.hadoopRDD.ignoreEmptySplits` como `true` por padrão. Como solução alternativa,

defina explicitamente `spark.hadoopRDD.ignoreEmptySplits` como `false`. A EMR versão 6.9.0 da Amazon corrige esse problema.

- Quando você usa o Spark com a formatação de localização de partições do Hive para ler dados no Amazon S3 e executa o Spark nas EMR versões 5.30.0 a 5.36.0 e 6.2.0 a 6.9.0 da Amazon, você pode encontrar um problema que impede seu cluster de ler os dados corretamente. Isso poderá acontecer se suas partições tiverem todas as características a seguir:
 - Duas ou mais partições são verificadas na mesma tabela.
 - Pelo menos um caminho de diretório de partição é um prefixo de pelo menos outro caminho de diretório de partição, por exemplo, `s3://bucket/table/p=a` é um prefixo de `s3://bucket/table/p=a b`.
 - O primeiro caractere que segue o prefixo no outro diretório de partição tem um valor UTF -8 menor que o / caractere (U+002F). Por exemplo, o caractere de espaço (U+0020) que ocorre entre a e b em `s3://bucket/table/p=a b` se enquadra nessa categoria. Observe que existem 14 outros caracteres que não são de controle: `!"#$%&'()*+,-.` Para obter mais informações, consulte a [tabela de codificação UTF -8 e caracteres Unicode](#).

Como solução alternativa para esse problema, defina a configuração `spark.sql.sources.fastS3PartitionDiscovery.enabled` como `false` na classificação `spark-defaults`.

- Com as EMR versões 5.36.0 e 6.6.0 a 6.9.0 da Amazon, `SecretAgent` os componentes do `RecordServer` serviço podem sofrer perda de dados de log devido a uma configuração incorreta do padrão de nome de arquivo nas propriedades do `Log4j2`. A configuração incorreta faz com que os componentes gerem somente um arquivo de log por dia. Quando a estratégia de rotação ocorre, ela substitui o arquivo existente em vez de gerar um novo arquivo de log, conforme esperado. Como solução alternativa, use uma ação de bootstrap para gerar arquivos de log a cada hora e acrescentar um número inteiro de incremento automático no nome do arquivo para lidar com a rotação.

Para as versões EMR 6.6.0 a 6.9.0 da Amazon, use a seguinte ação de bootstrap ao iniciar um cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Para o Amazon EMR 5.36.0, use a seguinte ação de bootstrap ao iniciar um cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

Para obter mais informações sobre o cronograma da versão, consulte o [log de alterações](#).

Versão 6.7.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 6.7.0 da Amazon. As alterações são referentes à versão 6.6.0.

Data da versão inicial: 15 de julho de 2022

Novos atributos

- A Amazon EMR agora oferece suporte ao Apache Spark 3.2.1, Apache Hive 3.1.3, HUDI 0.11, PrestoDB 0.272 e Trino 0.378.
- Suporta controles de acesso baseados em IAM Role e Lake Formation com EMR etapas (Spark, Hive) para Amazon em clusters. EMR EC2
- Ele é compatível com instruções de definição de dados do Apache Spark em clusters habilitados para Apache Ranger. Isso agora inclui suporte para aplicações do Trino lendo e gravando metadados do Apache Hive em clusters habilitados para Apache Ranger. Para obter mais informações, consulte [Habilitar a governança federada usando Trino e Apache Ranger na Amazon EMR](#).
- Quando você inicia um cluster com a versão de patch mais recente do Amazon EMR 5.36 ou superior, 6.6 ou superior ou 7.0 ou superior, a Amazon EMR usa a versão mais recente do Amazon Linux 2023 ou Amazon Linux 2 para a Amazon padrão. EMR AMI Para obter mais informações, consulte [Usando o Amazon Linux padrão AMI para a Amazon EMR](#).

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 709.1	4.14.348	23 de julho de 2024	Leste dos EUA (Norte da Virgínia), Leste dos

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			<p>EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), AWS GovCloud China (Pequim), China AWS GovCloud (Ningxia), Ásia Pacífico (Hyderabad), Oriente Médio (UAE), Europa (Espanha), Europa (Zurique)</p>

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 223.0	4.14.336	8 de março de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 131.0	4.14.336	14 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 124.0	4.14.336	7 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 109.0	4.14.334	24 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 218.0	4.14.330	2 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 206.0	4.14.330	22 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 116.0	4.14.328	11 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 101.0	4.14.327	16 de novembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 020.1	4.14.326	7 de novembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 012.1	4.14.326	26 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 926.0	4.14.322	19 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 906.0	4.14.322	4 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 822.0	4.14.322	30 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 808.0	4.14.320	24 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 727.0	4.14.320	14 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 719.0	4.14.320	2 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Oriente Médio (), Canadá (Central) UAE

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 628.0	4.14.318	12 de julho de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 612.0	4.14.314	23 de junho de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 504.1	4.14.313	16 de maio de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 418.0	4.14.311	3 de maio de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 404.1	4.14.311	18 de abril de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)
2.0.2023 404.0	4.14.311	10 de abril de 2023	Leste dos EUA (Norte da Virgínia), Europa (Paris)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 320.0	4.14.309	30 de março de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 307.0	4.14.305	15 de março de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 207.0	4.14.304	22 de fevereiro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 119.1	4.14.301	3 de fevereiro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 210.1	4.14.301	22 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 103.3	4.14.296	5 de dezembro de 2022	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 004.0	4.14.294	2 de novembro de 2022	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 912.1	4.14.291	7 de outubro de 2022	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)
2.0.2022 719.0	4.14.287	10 de agosto de 2022	us-west-1 , eu-west-3 , eu-north-1 , ap-south-1 , me-south-1

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 606.1	4.14.281	15 de julho de 2022	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

Problemas conhecidos

- Quando a Amazon EMR versão 6.5.0, 6.6.0 ou 6.7.0 lê as tabelas do Apache Phoenix por meio do shell do Apache Spark, a ocorre porque a Amazon usa um `NoSuchMethodError` erro. `EMR Hbase.compat.version` A EMR versão 6.8.0 da Amazon corrige esse problema.

- Quando você usa o conector do DynamoDB com o Spark nas versões 6.6.0, 6.7.0 e 6.8.0 da EMR Amazon, todas as leituras da sua tabela retornam um resultado vazio, mesmo que a divisão de entrada faça referência a dados não vazios. Isso ocorre porque o Spark 3.2.0 define `spark.hadoopRDD.ignoreEmptySplits` como `true` por padrão. Como solução alternativa, defina explicitamente `spark.hadoopRDD.ignoreEmptySplits` como `false`. A EMR versão 6.9.0 da Amazon corrige esse problema.
- Quando você usa o Spark com a formatação de localização de partições do Hive para ler dados no Amazon S3 e executa o Spark nas EMR versões 5.30.0 a 5.36.0 e 6.2.0 a 6.9.0 da Amazon, você pode encontrar um problema que impede seu cluster de ler os dados corretamente. Isso poderá acontecer se suas partições tiverem todas as características a seguir:
 - Duas ou mais partições são verificadas na mesma tabela.
 - Pelo menos um caminho de diretório de partição é um prefixo de pelo menos outro caminho de diretório de partição, por exemplo, `s3://bucket/table/p=a` é um prefixo de `s3://bucket/table/p=a b`.
 - O primeiro caractere que segue o prefixo no outro diretório de partição tem um valor UTF -8 menor que o / caractere (U+002F). Por exemplo, o caractere de espaço (U+0020) que ocorre entre a e b em `s3://bucket/table/p=a b` se enquadra nessa categoria. Observe que existem 14 outros caracteres que não são de controle: `!"#$%&'()*+,-.` Para obter mais informações, consulte a [tabela de codificação UTF -8 e caracteres Unicode](#).

Como solução alternativa para esse problema, defina a configuração `spark.sql.sources.fastS3PartitionDiscovery.enabled` como `false` na classificação `spark-defaults`.

- Com as EMR versões 5.36.0 e 6.6.0 a 6.9.0 da Amazon, `SecretAgent` os componentes do `RecordServer` serviço podem sofrer perda de dados de log devido a uma configuração incorreta do padrão de nome de arquivo nas propriedades do `Log4j2`. A configuração incorreta faz com que os componentes gerem somente um arquivo de log por dia. Quando a estratégia de rotação ocorre, ela substitui o arquivo existente em vez de gerar um novo arquivo de log, conforme esperado. Como solução alternativa, use uma ação de bootstrap para gerar arquivos de log a cada hora e acrescentar um número inteiro de incremento automático no nome do arquivo para lidar com a rotação.

Para as versões EMR 6.6.0 a 6.9.0 da Amazon, use a seguinte ação de bootstrap ao iniciar um cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Para o Amazon EMR 5.36.0, use a seguinte ação de bootstrap ao iniciar um cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

- O `GetClusterSessionCredentials` API não é compatível com clusters executados na Amazon EMR 6.7 ou inferior.

Versão 6.6.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 6.6.0 da Amazon. As alterações são referentes à versão 6.5.0.

Data da versão inicial: 9 de maio de 2022

Data da documentação atualizada: 15 de junho de 2022

Novos atributos

- O Amazon EMR 6.6 agora é compatível com Apache Spark 3.2, Apache Spark RAPIDS 22.02, CUDA 11, Apache Hudi 0.10.1, Apache Iceberg 0.13, Trino 0.367 e PrestoDB 0.267.
- Quando você inicia um cluster com a versão de patch mais recente do Amazon EMR 5.36 ou superior, 6.6 ou superior ou 7.0 ou superior, a Amazon EMR usa a versão mais recente do Amazon Linux 2023 ou Amazon Linux 2 para a Amazon padrão. EMR AMI Para obter mais informações, consulte [Usando o Amazon Linux padrão AMI para a Amazon EMR](#).

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 709.1	4.14.348	23 de julho de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
			<p>dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China AWS GovCloud (Ningxia)</p>

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 223.0	4.14.336	8 de março de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 131.0	4.14.336	14 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 124.0	4.14.336	7 de fevereiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2024 109.0	4.14.334	24 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 218.0	4.14.330	2 de janeiro de 2024	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), (Leste dos EUA), China AWS GovCloud (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 206.0	4.14.330	22 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 116.0	4.14.328	11 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 101.0	4.14.327	16 de novembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 020.1	4.14.326	7 de novembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 012.1	4.14.326	26 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 926.0	4.14.322	19 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka), Ásia-Pacífico (Osaka) Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central), (Oeste dos EUA), AWS GovCloud (Leste dos EUA), China (Pequim), China AWS GovCloud (Ningxia)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 906.0	4.14.322	4 de outubro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 822.0	4.14.322	30 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 808.0	4.14.320	24 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 727.0	4.14.320	14 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Frankfurt), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Canadá (Central)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 719.0	4.14.320	2 de agosto de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Europa (Estocolmo), Europa (Milão), Europa (Espanha), Europa (Frankfurt), Europa (Zurique), Europa (Irlanda), Europa (Londres), Europa (Paris), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad), Ásia-Pacífico (Hyderabad) Tóquio), Ásia-Pacífico (Seul), Ásia-Pacífico (Osaka), Ásia-Pacífico (Cingapura), Ásia-Pacífico (Sydney), Ásia-Pacífico (Jacarta), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein), Oriente Médio (), Oriente Médio (), Canadá (Central) UAE

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 628.0	4.14.318	12 de julho de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 612.0	4.14.314	23 de junho de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 504.1	4.14.313	16 de maio de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 418.0	4.14.311	3 de maio de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 404.1	4.14.311	18 de abril de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)
2.0.2023 404.0	4.14.311	10 de abril de 2023	Leste dos EUA (Norte da Virgínia), Europa (Paris)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 320.0	4.14.309	30 de março de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 307.0	4.14.305	15 de março de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 207.0	4.14.304	22 de fevereiro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2023 119.1	4.14.301	3 de fevereiro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 210.1	4.14.301	22 de dezembro de 2023	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 103.3	4.14.296	5 de dezembro de 2022	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 004.0	4.14.294	2 de novembro de 2022	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 912.1	4.14.291	7 de outubro de 2022	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)
2.0.2022 805.0	4.14.287	30 de agosto de 2022	us-west-1

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 719.0	4.14.287	10 de agosto de 2022	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 426.0	4.14.281	10 de junho de 2022	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

OsRelea Label (Versão Amazon Linux)	Versão do kernel do Amazon Linux	Data disponível	Regiões compatíveis
2.0.2022 406.1	4.14.275	2 de maio de 2022	Leste dos EUA (Norte da Virgínia), Leste dos EUA (Ohio), Oeste dos EUA (Norte da Califórnia), Oeste dos EUA (Oregon), Canadá (Central), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (Paris), Europa (Frankfurt), Europa (Milão), Ásia-Pacífico (Hong Kong), Ásia-Pacífico (Mumbai), Ásia-Pacífico (Jacarta), Ásia-Pacífico (Tóquio), Ásia-Pacífico (Seul), Asia Pacific (Osaka), Ásia-Pacífico (Singapura), Ásia-Pacífico (Sydney), África (Cidade do Cabo), América do Sul (São Paulo), Oriente Médio (Bahrein)

- [Com o Amazon EMR 6.6 e versões posteriores, os aplicativos que usam o Log4j 1.x e o Log4j 2.x são atualizados para usar o Log4j 1.2.17 \(ou superior\) e o Log4j 2.17.1 \(ou superior\), respectivamente, e não exigem o uso das ações de bootstrap fornecidas para mitigar os problemas.](#) CVE
- [Escalabilidade gerenciada] Otimização de escalabilidade gerenciada de dados do Spark shuffle
 - Para EMR as versões 5.34.0 e posteriores da Amazon e as EMR versões 6.4.0 e posteriores, a escalabilidade gerenciada agora reconhece os dados do Spark shuffle (dados que o Spark

redistribui entre partições para realizar operações específicas). Para obter mais informações sobre operações aleatórias, consulte Como [usar a escalabilidade EMR gerenciada na Amazon no Guia de EMR Gerenciamento EMR da Amazon](#) e no Guia de Programação do [Spark](#).

- A partir do Amazon EMR 5.32.0 e 6.5.0, o dimensionamento dinâmico do executor para o Apache Spark está habilitado por padrão. Para ativar ou desativar esse atributo, você pode usar o parâmetro de configuração `spark.yarn.heterogeneousExecutors.enabled`.

Alterações, melhorias e problemas resolvidos

- A Amazon EMR reduz o tempo de inicialização do cluster em até 80 segundos, em média, para clusters que usam a AMI opção EMR padrão e instalam apenas aplicativos comuns, como Apache Hadoop, Apache Spark e Apache Hive.

Problemas conhecidos

- Quando a Amazon EMR versão 6.5.0, 6.6.0 ou 6.7.0 lê as tabelas do Apache Phoenix por meio do shell do Apache Spark, a ocorre porque a Amazon usa um `NoSuchMethodError` erro. `EMR Hbase.compat.version` A EMR versão 6.8.0 da Amazon corrige esse problema.
- Quando você usa o conector do DynamoDB com o Spark nas versões 6.6.0, 6.7.0 e 6.8.0 da EMR Amazon, todas as leituras da sua tabela retornam um resultado vazio, mesmo que a divisão de entrada faça referência a dados não vazios. Isso ocorre porque o Spark 3.2.0 define `spark.hadoopRDD.ignoreEmptySplits` como `true` por padrão. Como solução alternativa, defina explicitamente `spark.hadoopRDD.ignoreEmptySplits` como `false`. A EMR versão 6.9.0 da Amazon corrige esse problema.
- Em clusters de longa duração do Trino, o Amazon EMR 6.6.0 habilita os parâmetros de registro da coleta de lixo no `jvm.config` do Trino para obter melhores insights dos registros da coleta de lixo. Essa alteração anexa muitos logs da coleta de resíduos ao arquivo `launcher.log` (`/var/log/trino/launcher.log`). Se você estiver executando clusters Trino no Amazon EMR 6.6.0, poderá encontrar nós sem espaço em disco depois que o cluster estiver em execução por alguns dias devido aos registros anexados.

A solução alternativa para esse problema é executar o script abaixo como uma ação de bootstrap para desativar os parâmetros de registro da coleta de lixo no `jvm.config` ao criar ou clonar o cluster para o 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"
```

- Quando você usa o Spark com a formatação de localização de partições do Hive para ler dados no Amazon S3 e executa o Spark nas EMR versões 5.30.0 a 5.36.0 e 6.2.0 a 6.9.0 da Amazon, você pode encontrar um problema que impede seu cluster de ler os dados corretamente. Isso poderá acontecer se suas partições tiverem todas as características a seguir:
 - Duas ou mais partições são verificadas na mesma tabela.
 - Pelo menos um caminho de diretório de partição é um prefixo de pelo menos outro caminho de diretório de partição, por exemplo, s3://bucket/table/p=a é um prefixo de s3://bucket/table/p=a b.
 - O primeiro caractere que segue o prefixo no outro diretório de partição tem um valor UTF -8 menor que o / caractere (U+002F). Por exemplo, o caractere de espaço (U+0020) que ocorre entre a e b em s3://bucket/table/p=a b se enquadra nessa categoria. Observe que existem 14 outros caracteres que não são de controle: !"#&'()*+,-. Para obter mais informações, consulte a [tabela de codificação UTF -8 e caracteres Unicode](#).

Como solução alternativa para esse problema, defina a configuração

`spark.sql.sources.fastS3PartitionDiscovery.enabled` como `false` na classificação `spark-defaults`.

- Com as EMR versões 5.36.0 e 6.6.0 a 6.9.0 da Amazon, `SecretAgent` os componentes do `RecordServer` serviço podem sofrer perda de dados de log devido a uma configuração incorreta do padrão de nome de arquivo nas propriedades do `Log4j2`. A configuração incorreta faz com que os componentes gerem somente um arquivo de log por dia. Quando a estratégia de rotação ocorre, ela substitui o arquivo existente em vez de gerar um novo arquivo de log, conforme esperado. Como solução alternativa, use uma ação de bootstrap para gerar arquivos de log a cada hora e acrescentar um número inteiro de incremento automático no nome do arquivo para lidar com a rotação.

Para as versões EMR 6.6.0 a 6.9.0 da Amazon, use a seguinte ação de bootstrap ao iniciar um cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-6x/replace-puppet.sh,Args=[]"
```

Para o Amazon EMR 5.36.0, use a seguinte ação de bootstrap ao iniciar um cluster.

```
--bootstrap-actions "Path=s3://emr-data-access-control-us-east-1/customer-bootstrap-actions/log-rotation-emr-5x/replace-puppet.sh,Args=[]"
```

Versão 5.35.0

Esta é a nota de EMR lançamento da Amazon versão 5.35.0.

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.35.0 da Amazon. As alterações são referentes à versão 5.34.0.

Data da versão inicial: 30 de março de 2022

Novos atributos

- Os aplicativos da Amazon EMR versão 5.35 que usam o Log4j 1.x e o Log4j 2.x são atualizados para usar o Log4j 1.2.17 (ou superior) e o Log4j 2.17.1 (ou superior), respectivamente, e não exigem o uso de ações de bootstrap para mitigar os problemas nas versões anteriores. CVE Consulte [Abordagem para CVE mitigar -2021-44228](#).

Alterações, melhorias e problemas resolvidos

Alterações no Flink

Alterar tipo	Descrição
Atualizações	<ul style="list-style-type: none"> Atualizar a versão do flink para 1.14.2. log4j atualizado para 2.17.1.

Alterações no Hadoop

Alterar tipo	Descrição
Backports de código aberto do Hadoop desde 5.34.0 EMR	<ul style="list-style-type: none"> YARN-10438: Manipulação de nulo container Id em C # () lientRMService getContainerReport YARN-7266: Tópicos do manipulador de eventos do Timeline Server bloqueados

Alterar tipo	Descrição
	<ul style="list-style-type: none">• YARN-10438: ATS 1.5 falha ao iniciar se os RollingLevelDb arquivos estiverem corrompidos ou ausentes• HADOOP-13500: Sincronizando a iteração do objeto de propriedades de configuração• YARN-10651: CapacityScheduler caiu com o in. NPE AbstractYarnScheduler updateNodeResource()• HDFS-12221: Substituir xerces em XmlEditsVisitor• HDFS-16410: Análise insegura de Xml em OfflineEditsXMLLoader
Alterações e correções no Hadoop	<ul style="list-style-type: none">• O Tomcat usado em KMS e o HTTPFs foi atualizado para 8.5.75• Na FileSystemOptimizedCommitter V2, o marcador de sucesso foi escrito no caminho de commitJob saída definido durante a criação do committer. Como commitJob os caminhos de saída no nível da tarefa podem ser diferentes, o caminho foi corrigido para usar aquele definido nos arquivos de manifesto. Para trabalhos do Hive, isso faz com que o marcador de sucesso seja gravado corretamente ao executar operações como partição dinâmica ou UNION ALL

Alterações no Hive

Alterar tipo	Descrição
<p>O Hive foi atualizado para a versão 2.3.9 de código aberto, incluindo essas correções JIRA</p>	<ul style="list-style-type: none"> • HIVE-17155: findConfFile () em HiveConf .java tem alguns problemas com o caminho de configuração • HIVE-24797: Desative a validação de valores padrão ao analisar esquemas Avro • HIVE-21563: Melhore o desempenho do getEmptyTable Table# desativando Once registerAllFunctions • HIVE-18147: Os testes podem falhar com java.net. BindException: Endereço já em uso • HIVE-24608: Volte para get_table no cliente para Hive 2.3.x HMS • HIVE-21200: Vetorização - coluna de data lançando java.lang. UnsupportedOperationException para parquet • HIVE-19228: Remover o uso do commons-httpclient 3.x
<p>Backports de código aberto do Hive desde 5.34.0 EMR</p>	<ul style="list-style-type: none"> • HIVE-19990: Falha na consulta com intervalo literal na condição de junção • HIVE-25824: Atualize o branch-2.3 para o log4j 2.17.0 • TEZ-4062: O agendamento de tentativas especulativas deve ser abortado quando a tarefa for concluída • TEZ-4108: NullPointerException durante a execução especulativa, condição de corrida • TEZ-3918: A configuração de tez.task.log.level não funciona
<p>Atualizações e correções do Hive</p>	<ul style="list-style-type: none"> • Atualizar a versão Log4j para 2.17.1 • Atualize a ORC versão para 1.4.3

Alterar tipo	Descrição
	<ul style="list-style-type: none"> Corrigido o impasse devido à discussão de penalidades em ShuffleScheduler
Novos atributos	<ul style="list-style-type: none"> Adicionado atributo para imprimir consultas do Hive em logs AM. Ela fica desabilitada por padrão. Sinalizador/Conf.: <code>tez.am.emr.print.hive.query.in.log</code>. Status (padrão):FALSE.

Alterações no Oozie

Alterar tipo	Descrição
Backports de código aberto Oozie desde 5.34.0 EMR	<ul style="list-style-type: none"> OOZIE-3652: O lançador Oozie deve tentar novamente a listagem de diretórios quando ocorrer <code>NoSuchFileException</code>

Alterações no Pig

Alterar tipo	Descrição
Atualizações	<ul style="list-style-type: none"> log4j atualizado para 1.2.17.

Problemas conhecidos

- Quando você usa o Spark com a formatação de localização de partições do Hive para ler dados no Amazon S3 e executa o Spark nas EMR versões 5.30.0 a 5.36.0 e 6.2.0 a 6.9.0 da Amazon, você pode encontrar um problema que impede seu cluster de ler os dados corretamente. Isso poderá acontecer se suas partições tiverem todas as características a seguir:
 - Duas ou mais partições são verificadas na mesma tabela.
 - Pelo menos um caminho de diretório de partição é um prefixo de pelo menos outro caminho de diretório de partição, por exemplo, `s3://bucket/table/p=a` é um prefixo de `s3://bucket/table/p=a b`.

- O primeiro caractere que segue o prefixo no outro diretório de partição tem um valor UTF -8 menor que o / caractere (U+002F). Por exemplo, o caractere de espaço (U+0020) que ocorre entre a e b em `s3://bucket/table/p=a b` se enquadra nessa categoria. Observe que existem 14 outros caracteres que não são de controle: `!"#$%&'()*+,-.` Para obter mais informações, consulte a [tabela de codificação UTF -8 e caracteres Unicode](#).

Como solução alternativa para esse problema, defina a configuração `spark.sql.sources.fastS3PartitionDiscovery.enabled` como `false` na classificação `spark-defaults`.

Versão 5.34.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.34.0 da Amazon. As alterações são referentes à versão 5.33.1.

Data da versão inicial: 20 de janeiro de 2022

Data da versão atualizada: 21 de março de 2022

Novos atributos

- [Escalabilidade gerenciada] Otimização de escalabilidade gerenciada de dados do Spark shuffle - Para EMR as versões 5.34.0 e posteriores da Amazon e as EMR versões 6.4.0 e posteriores, a escalabilidade gerenciada agora reconhece os dados do Spark shuffle (dados que o Spark redistribui entre partições para realizar operações específicas). Para obter mais informações sobre operações aleatórias, consulte Como [usar a escalabilidade EMR gerenciada na Amazon no Guia de EMR Gerenciamento EMR da Amazon](#) e no Guia de Programação do [Spark](#).
- [Hudi] Melhorias para simplificar a configuração do Hudi. Desabilitado o controle de simultaneidade otimista por padrão.

Alterações, melhorias e problemas resolvidos

- Esta é uma versão para corrigir problemas com o Amazon EMR Scaling quando ele falha na escalabilidade ou redução de escala de um cluster com sucesso ou causa falhas no aplicativo.
- Anteriormente, a reinicialização manual do gerenciador de recursos em um cluster multimestre fazia com que os daemons EMR no cluster da Amazon, como o Zookeeper, recarregassem todos os nós anteriormente desativados ou perdidos no arquivo `znode` do Zookeeper. Isso fez com que os limites padrão fossem excedidos em determinadas situações. A Amazon EMR agora remove

os registros de nós desativados ou perdidos há mais de uma hora do arquivo do Zookeeper e os limites internos foram aumentados.

- Corrigido um problema em que as solicitações de escalabilidade falhavam em um cluster grande e altamente utilizado quando os daemons no cluster da EMR Amazon estavam executando atividades de verificação de integridade, como coletar o estado do nó e o estado do YARN nó. HDFS Isso estava acontecendo porque os daemons no cluster não conseguiam comunicar os dados do status de saúde de um nó aos componentes internos da Amazon. EMR
- Daemons EMR no cluster aprimorados para rastrear corretamente os estados dos nós quando os endereços IP são reutilizados para melhorar a confiabilidade durante as operações de escalabilidade.
- [SPARK-29683](#). Corrigido um problema em que ocorriam falhas de trabalho durante a redução da escala verticalmente do cluster, pois o Spark presumia que todos os nós disponíveis estavam na lista de negação.
- [YARN-9011](#). Corrigido um problema em que falhas de trabalho ocorriam devido a uma condição de corrida no YARN descomissionamento quando o cluster tentava aumentar ou diminuir a escala.
- Foi corrigido o problema com falhas de etapas ou tarefas durante a escalabilidade do cluster, garantindo que os estados dos nós fossem sempre consistentes entre os EMR daemons no cluster da Amazon e/. YARN HDFS
- Corrigido um problema em que operações de cluster, como redução de escala e envio de etapas, falhavam em EMR clusters da Amazon habilitados com a autenticação Kerberos. Isso ocorreu porque o daemon no EMR cluster da Amazon não renovou o tíquete Kerberos, que é necessário para se comunicar com segurança com/em execução no nó primário. HDFS YARN
- Zeppelin atualizado para a versão 0.10.0.
- Livy Fix: atualizado para 0.7.1
- Melhoria do desempenho do Spark - executores heterogêneos são desativados quando determinados valores de configuração do Spark são substituídos na versão 5.34.0. EMR
- Os servidores Web HDFS e HTTPFs estão desativados por padrão. Você pode reativar a Web HDFS usando a configuração do Hadoop, `dfs.webhdfs.enabled` O servidor HttpFS pode ser iniciado usando `sudo systemctl start hadoop-httpfs`.

Problemas conhecidos

- O recurso Amazon EMR Notebooks usado com a representação de usuário do Livy não funciona porque o HTTPFs está desativado por padrão. Nesse caso, o EMR notebook não pode se conectar

ao cluster que tem a representação Livy ativada. A solução alternativa é iniciar o servidor HTTPFs antes de conectar o EMR notebook ao cluster usando o `sudo systemctl start hadoop-httpfs`

- As consultas Hue não funcionam no Amazon EMR 6.4.0 porque o servidor Apache Hadoop HTTPFs está desativado por padrão. Para usar o Hue na Amazon EMR 6.4.0, inicie manualmente o servidor HTTPFs no nó EMR primário da Amazon usando `sudo systemctl start hadoop-httpfs` ou use [uma etapa da Amazon](#). EMR
- O recurso Amazon EMR Notebooks usado com a representação de usuário do Livy não funciona porque o HTTPFs está desativado por padrão. Nesse caso, o EMR notebook não pode se conectar ao cluster que tem a representação Livy ativada. A solução alternativa é iniciar o servidor HTTPFs antes de conectar o EMR notebook ao cluster usando o `sudo systemctl start hadoop-httpfs`
- Quando você usa o Spark com a formatação de localização de partições do Hive para ler dados no Amazon S3 e executa o Spark nas EMR versões 5.30.0 a 5.36.0 e 6.2.0 a 6.9.0 da Amazon, você pode encontrar um problema que impede seu cluster de ler os dados corretamente. Isso poderá acontecer se suas partições tiverem todas as características a seguir:
 - Duas ou mais partições são verificadas na mesma tabela.
 - Pelo menos um caminho de diretório de partição é um prefixo de pelo menos outro caminho de diretório de partição, por exemplo, `s3://bucket/table/p=a` é um prefixo de `s3://bucket/table/p=a b`.
 - O primeiro caractere que segue o prefixo no outro diretório de partição tem um valor UTF-8 menor que o `/` caractere (U+002F). Por exemplo, o caractere de espaço (U+0020) que ocorre entre `a` e `b` em `s3://bucket/table/p=a b` se enquadra nessa categoria. Observe que existem 14 outros caracteres que não são de controle: `!"#$%&'()*+,-.` Para obter mais informações, consulte a [tabela de codificação UTF-8 e caracteres Unicode](#).

Como solução alternativa para esse problema, defina a configuração `spark.sql.sources.fastS3PartitionDiscovery.enabled` como `false` na classificação `spark-defaults`.

Versão 6.5.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 6.5.0 da Amazon. As alterações são referentes à versão 6.4.0.

Data da versão inicial: 20 de janeiro de 2022

Data da versão atualizada: 21 de março de 2022

Novos atributos

- [Escalabilidade gerenciada] Otimização de escalabilidade gerenciada de dados do Spark shuffle - Para EMR as versões 5.34.0 e posteriores da Amazon e as EMR versões 6.4.0 e posteriores, a escalabilidade gerenciada agora reconhece os dados do Spark shuffle (dados que o Spark redistribui entre partições para realizar operações específicas). Para obter mais informações sobre operações aleatórias, consulte Como [usar a escalabilidade EMR gerenciada na Amazon no Guia de EMR Gerenciamento EMR da Amazon](#) e no Guia de Programação do [Spark](#).
- A partir do Amazon EMR 5.32.0 e 6.5.0, o dimensionamento dinâmico do executor para o Apache Spark está habilitado por padrão. Para ativar ou desativar esse atributo, você pode usar o parâmetro de configuração `spark.yarn.heterogeneousExecutors.enabled`.
- Suporte para o formato de tabela aberta Apache Iceberg para conjuntos de dados analíticos imensos.
- Support para ranger-trino-plugin 2.0.1-amzn-1
- Suporte para toree 0.5.0

Alterações, melhorias e problemas resolvidos

- A versão de lançamento do Amazon EMR 6.5 agora oferece suporte ao Apache Iceberg 0.12.0 e fornece melhorias no tempo de execução com o Amazon Runtime para Apache Spark, o Amazon EMR Runtime EMR for Presto e o Amazon Runtime para Apache Hive. EMR
- O [Apache Iceberg](#) é um formato de tabela aberta para grandes conjuntos de dados no Amazon S3 e fornece desempenho rápido de consultas em tabelas grandes, confirmações atômicas, gravações simultâneas e evolução de tabelas compatíveis. SQL Com a EMR versão 6.5, você pode usar o Apache Spark 3.1.2 com o formato de tabela Iceberg.
- O Apache Hudi 0.9 adiciona Spark e suporte SQLDDL. DML Isso permite que você crie e altere tabelas Hudi usando apenas declarações. SQL O Apache Hudi 0.9 também inclui melhorias na performance do lado da consulta e do lado do gravador.
- O Amazon EMR Runtime para Apache Hive melhora o desempenho do Apache Hive no Amazon S3 ao remover as operações de renomeação durante as operações de preparação e melhora o desempenho dos comandos metastore check () usados para reparar tabelas. MSCK

Problemas conhecidos

- Quando a Amazon EMR versão 6.5.0, 6.6.0 ou 6.7.0 lê as tabelas do Apache Phoenix por meio do shell do Apache Spark, a ocorre porque a Amazon usa um `NoSuchMethodError` erro. EMR `Hbase.compat.version` A EMR versão 6.8.0 da Amazon corrige esse problema.
- Os clusters do pacote do Hbase em alta disponibilidade (HA) apresentam falha no provisionamento com o tamanho de volume e o tipo de instância padrão. A solução alternativa para esse problema é aumentar o tamanho do volume raiz.
- Para usar as ações do Spark com o Apache Oozie, você deve adicionar a seguinte configuração ao seu arquivo `workflow.xml` do Oozie. Caso contrário, várias bibliotecas críticas, como a Hadoop e a Hadoop, EMRFS estarão ausentes do classpath dos executores do Spark que o Oozie lança.

```
<spark-opts>--conf spark.yarn.populateHadoopClasspath=true</spark-opts>
```

- Quando você usa o Spark com a formatação de localização de partições do Hive para ler dados no Amazon S3 e executa o Spark nas EMR versões 5.30.0 a 5.36.0 e 6.2.0 a 6.9.0 da Amazon, você pode encontrar um problema que impede seu cluster de ler os dados corretamente. Isso poderá acontecer se suas partições tiverem todas as características a seguir:
 - Duas ou mais partições são verificadas na mesma tabela.
 - Pelo menos um caminho de diretório de partição é um prefixo de pelo menos outro caminho de diretório de partição, por exemplo, `s3://bucket/table/p=a` é um prefixo de `s3://bucket/table/p=a b`.
 - O primeiro caractere que segue o prefixo no outro diretório de partição tem um valor UTF -8 menor que o / caractere (U+002F). Por exemplo, o caractere de espaço (U+0020) que ocorre entre a e b em `s3://bucket/table/p=a b` se enquadra nessa categoria. Observe que existem 14 outros caracteres que não são de controle: `!"#$%&'()*+,-.` Para obter mais informações, consulte a [tabela de codificação UTF -8 e caracteres Unicode](#).

Como solução alternativa para esse problema, defina a configuração `spark.sql.sources.fastS3PartitionDiscovery.enabled` como `false` na classificação `spark-defaults`.

Versão 6.4.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 6.4.0 da Amazon. As alterações são referentes à versão 6.3.0.

Data da versão inicial: 20 de setembro de 2021

Data da versão atualizada: 21 de março de 2022

Aplicações compatíveis

- AWS SDK for Java versão 1.12.31
- CloudWatch Sink versão 2.2.0
- Conector do DynamoDB versão 4.16.0
- EMRFS versão 2.47.0
- Amazon EMR Goodies versão 3.2.0
- Amazon EMR Kinesis Connector versão 3.5.0
- Amazon EMR Record Server versão 2.1.0
- Amazon EMR Scripts versão 2.5.0
- Flink versão 1.13.1
- Ganglia versão 3.7.2
- AWS Glue Hive Metastore Client versão 3.3.0
- Hadoop versão 3.2.1-amzn-4
- HBase versão 2.4.4-amzn-0
- HBase-operador-tools 1.1.0
- HCatalog versão 3.1.2-amzn-5
- Hive versão 3.1.2-amzn-5
- Hudi versão 0.8.0-amzn-0
- Hue versão 4.9.0
- JDK Versão Java Corretto-8.302.08.1 (compilação 1.8.0_302-b08)
- JupyterHub versão 1.4.1
- Livy versão 0.7.1-incubating
- MXNet versão 1.8.0
- Oozie versão 5.2.1

- Phoenix versão 5.1.2
- Pig versão 0.17.0
- Presto versão 0.254.1-amzn-0
- Trino versão 359
- Apache Ranger KMS (criptografia transparente multi-master) versão 2.0.0
- ranger-plugins 2.0.1-amzn-0
- ranger-s3-plugin 1.2.0
- SageMaker Spark SDK versão 1.4.1
- Scala versão 2.12.10 (VM de servidor aberto de JDK 64 bits, Java 1.8.0_282)
- Spark versão 3.1.2-amzn-0
- spark-rapids 0.4.1
- Sqoop versão 1.4.7
- TensorFlow versão 2.4.1
- Tez versão 0.9.2
- Zeppelin versão 0.9.0
- Zookeeper versão 3.5.7
- Conectores e drivers: DynamoDB Connector 4.16.0

Novos atributos

- [Escalabilidade gerenciada] Otimização de escalabilidade gerenciada de dados do Spark shuffle - Para EMR as versões 5.34.0 e posteriores da Amazon e as EMR versões 6.4.0 e posteriores, a escalabilidade gerenciada agora reconhece os dados do Spark shuffle (dados que o Spark redistribui entre partições para realizar operações específicas). Para obter mais informações sobre operações aleatórias, consulte Como [usar a escalabilidade EMR gerenciada na Amazon no Guia de EMR Gerenciamento EMR da Amazon](#) e no Guia de Programação do [Spark](#).
- Em clusters EMR Amazon habilitados para Apache Ranger, você pode usar o Apache SQL Spark para inserir dados ou atualizar as tabelas de metastore do Apache Hive usando, e. INSERT INTO INSERT OVERWRITE ALTER TABLE Ao usar ALTER TABLE com o SparkSQL, um local de partição deve ser o diretório filho de um local de tabela. Atualmente, EMR a Amazon não suporta a inserção de dados em uma partição em que a localização da partição seja diferente da localização da tabela.
- Presto SQL foi [renomeado para Trino](#).

- Hive: a execução de SELECT consultas simples com LIMIT cláusula é acelerada interrompendo a execução da consulta assim que o número de registros mencionados na LIMIT cláusula for obtido. SELECTConsultas simples são consultas que não têm a cláusula GROUP BY/ORDERby ou consultas que não têm um estágio redutor. Por exemplo, `SELECT * from <TABLE> WHERE <Condition> LIMIT <Number>`.

Controle de simultaneidade do Hudi

- O Hudi agora oferece suporte ao Optimistic Concurrency Control (OCC), que pode ser aproveitado com operações de gravação, como UPSERT e INSERT para permitir alterações de vários gravadores na mesma tabela Hudi. Isso é no nível do arquivoOCC, então quaisquer dois commits (ou escritores) podem gravar na mesma tabela, se suas alterações não entrarem em conflito. Para obter mais informações, consulte o [Controle de simultaneidade do Hudi](#).
- EMROs clusters da Amazon têm o Zookeeper instalado, que pode ser usado como provedor de bloqueio para. OCC Para facilitar o uso desse recurso, os EMR clusters da Amazon têm as seguintes propriedades pré-configuradas:

```
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
```

Para habilitarOCC, você precisa configurar as seguintes propriedades com suas opções de trabalho Hudi ou no nível do cluster usando as configurações da 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)
```

Monitoramento Hudi: CloudWatch integração com a Amazon para reportar Hudi Metrics

- A Amazon EMR oferece suporte à publicação de Hudi Metrics na Amazon CloudWatch. Isso é habilitado com a definição das seguintes configurações necessárias:

```
hoodie.metrics.on=true
```

```
hoodie.metrics.reporter.type=CLOUDWATCH
```

- A seguir, são mostradas as configurações opcionais do Hudi que você pode alterar:

Configuração	Descrição	Valor
hoodie.metrics.cloudwatch.report.period.seconds	Frequência (em segundos) na qual reportar métricas para a Amazon CloudWatch	O valor padrão é 60s, o que é adequado para a resolução padrão de um minuto oferecida pela Amazon CloudWatch
hoodie.metrics.cloudwatch.metric.prefix	Prefixo a ser adicionado ao nome de cada métrica	O valor padrão está vazio (sem prefixo)
hoodie.metrics.cloudwatch.namespace	CloudWatch Namespace da Amazon sob o qual as métricas são publicadas	O valor padrão é Hudi
hoodie.metrics.cloudwatch.maxDatumsPerSolicitação	Número máximo de dados a serem incluídos em uma solicitação à Amazon CloudWatch	O valor padrão é 20, que é o mesmo que o CloudWatch padrão da Amazon

Suporte e melhorias nas configurações do Amazon EMR Hudi

- Agora, os clientes podem aproveitar o recurso de EMR configurações API e reconfiguração para configurar as configurações do Hudi no nível do cluster. Um novo suporte à configuração baseado em arquivos foi introduzido por meio de `/etc/hudi/conf/hudi-defaults.conf` nos moldes de outras aplicações, como Spark, Hive etc. EMR configura alguns padrões para melhorar a experiência do usuário:

— `hoodie.datasource.hive_sync.jdbcurl` está configurado para o servidor Hive do cluster URL e não precisa mais ser especificado. Isso é particularmente útil ao executar um trabalho no modo de cluster do Spark, onde anteriormente era necessário especificar o IP EMR principal da Amazon.

— configurações HBase específicas, que são úteis para usar o HBase índice com o Hudi.

- Configuração específica do provedor de bloqueio do Zookeeper, conforme discutido em Controle de simultaneidade, o que facilita o uso do Controle de Concorrência Otimista (). OCC
- Alterações adicionais foram introduzidas para reduzir o número de configurações que você precisa passar e inferir automaticamente sempre que possível:
 - A palavra-chave `partitionBy` pode ser usada para especificar a coluna de partição.
 - Ao habilitar o Hive Sync, não é mais obrigatório passar `HIVE_TABLE_OPT_KEY`, `HIVE_PARTITION_FIELDS_OPT_KEY`, `HIVE_PARTITION_EXTRACTOR_CLASS_OPT_KEY`. Esses valores podem ser deduzidos com base no nome da tabela Hudi e no campo de partição.
 - não é obrigatório passar `KEYGENERATOR_CLASS_OPT_KEY`, que pode ser inferido com base em casos mais simples de `SimpleKeyGenerator` e `ComplexKeyGenerator`.

Advertências do Hudi

- O Hudi não permite execução vetorizada no Hive de tabelas Merge on Read (MoR - Mesclar na leitura) e Bootstrap. Por exemplo, `count(*)` apresenta falha com a tabela do Hudi em tempo real quando `hive.vectorized.execution.enabled` está definido como verdadeiro. Como solução alternativa, você pode desabilitar a leitura vetorizada configurando `hive.vectorized.execution.enabled` como `false`.
- O suporte a vários gravadores não é compatível com o atributo de bootstrap do Hudi.
- O Flink Streamer e o Flink SQL são recursos experimentais nesta versão. Esses atributos não são recomendados para uso em implantações de produção.

Alterações, melhorias e problemas resolvidos

Esta é uma versão para corrigir problemas com o Amazon EMR Scaling quando ele falha na escalabilidade ou redução de escala de um cluster com sucesso ou causa falhas no aplicativo.

- Anteriormente, a reinicialização manual do gerenciador de recursos em um cluster multimestre fazia com que os daemons EMR no cluster da Amazon, como o Zookeeper, recarregassem todos os nós anteriormente desativados ou perdidos no arquivo `znode` do Zookeeper. Isso fez com que os limites padrão fossem excedidos em determinadas situações. A Amazon EMR agora remove os registros de nós desativados ou perdidos há mais de uma hora do arquivo do Zookeeper e os limites internos foram aumentados.

- Corrigido um problema em que as solicitações de escalabilidade falhavam em um cluster grande e altamente utilizado quando os daemons no cluster da EMR Amazon estavam executando atividades de verificação de integridade, como coletar o estado do nó e o estado do YARN nó. HDFS Isso estava acontecendo porque os daemons no cluster não conseguiam comunicar os dados do status de saúde de um nó aos componentes internos da Amazon. EMR
- Daemons EMR no cluster aprimorados para rastrear corretamente os estados dos nós quando os endereços IP são reutilizados para melhorar a confiabilidade durante as operações de escalabilidade.
- [SPARK-29683](#). Corrigido um problema em que ocorriam falhas de trabalho durante a redução da escala verticalmente do cluster, pois o Spark presumia que todos os nós disponíveis estavam na lista de negação.
- [YARN-9011](#). Corrigido um problema em que falhas de trabalho ocorriam devido a uma condição de corrida no YARN descomissionamento quando o cluster tentava aumentar ou diminuir a escala.
- Foi corrigido o problema com falhas de etapas ou tarefas durante a escalabilidade do cluster, garantindo que os estados dos nós fossem sempre consistentes entre os EMR daemons no cluster da Amazon e/. YARN HDFS
- Corrigido um problema em que operações de cluster, como redução de escala e envio de etapas, falhavam em EMR clusters da Amazon habilitados com a autenticação Kerberos. Isso ocorreu porque o daemon no EMR cluster da Amazon não renovou o tíquete Kerberos, que é necessário para se comunicar com segurança com/em execução no nó primário. HDFS YARN
- Configurando um cluster para corrigir problemas de desempenho do Apache YARN Timeline Server versões 1 e 1.5

As versões 1 e 1.5 do Apache YARN Timeline Server podem causar problemas de desempenho com EMR clusters grandes e muito ativos, especialmente `com.yarn.resourcemanager.system-metrics-publisher.enabled=true`, que é a configuração padrão na Amazon. EMR Um YARN Timeline Server v2 de código aberto resolve o problema de desempenho relacionado à escalabilidade do YARN Timeline Server.

Outras soluções alternativas para esse problema incluem:

- Configurando `yarn.resourcemanager.system-metrics-publisher.enabled=false` em `yarn-site.xml`.
- Habilitar a correção para esse problema na criação de um cluster, conforme descrito abaixo.

As seguintes EMR versões da Amazon contêm uma correção para esse problema de desempenho YARN do Timeline Server.

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

Para habilitar a correção em qualquer uma das EMR versões da Amazon especificadas acima, defina essas propriedades `true` em um JSON arquivo de configurações que é passado usando o [parâmetro de aws emr create-cluster comando](#): `--configurations file:///./configurations.json`. Ou habilite a correção usando a [interface do usuário do console de reconfiguração](#).

Exemplo de conteúdo do arquivo `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": []
}
]
```

- Os servidores Web HDFS e HTTPFs estão desativados por padrão. Você pode reativar a Web HDFS usando a configuração do Hadoop, `dfs.webhdfs.enabled`. O servidor HttpFS pode ser iniciado usando `sudo systemctl start hadoop-httpfs`.
- HTTPS agora está habilitado por padrão para os repositórios Amazon Linux. Se você estiver usando uma VPCE política do Amazon S3 para restringir o acesso a buckets específicos, deverá adicionar o novo bucket Amazon Linux ARN `arn:aws:s3:::amazonlinux-2-repos-$region/*` à sua política (`$region` substitua pela região onde está o endpoint). Para obter mais informações, consulte esse tópico nos fóruns de AWS discussão. [Anúncio: o Amazon Linux 2 agora oferece suporte à capacidade de uso HTTPS ao se conectar a repositórios de pacotes](#).
- Hive: o desempenho da consulta de gravação foi aprimorado ao permitir o uso de um diretório de rascunho ativado HDFS para o último trabalho. Os dados temporários do trabalho final são gravados no Amazon S3, e o desempenho é aprimorado porque os dados são movidos HDFS para o local final da tabela (Amazon S3) em vez de entre dispositivos Amazon S3. HDFS
- Hive: melhoria do tempo de compilação de consultas em até 2,5 vezes com a remoção de partições de metastores do Glue.

- Por padrão, quando UDFs os integrados são passados pelo Hive para o Hive Metastore Server, somente um subconjunto desses incorporados é UDFs passado para o Glue Metastore, já que o Glue suporta apenas operadores de expressão limitados. Se você definir `hive.glue.partition.pruning.client=true`, toda a remoção de partições ocorrerá no lado do cliente. Se você definir `hive.glue.partition.pruning.server=true`, toda a remoção de partições ocorrerá no lado do servidor.

Problemas conhecidos

- As consultas Hue não funcionam no Amazon EMR 6.4.0 porque o servidor Apache Hadoop HTTPFs está desativado por padrão. Para usar o Hue na Amazon EMR 6.4.0, inicie manualmente o servidor HTTPFs no nó EMR primário da Amazon usando `sudo systemctl start hadoop-httpfs` ou use [uma etapa da Amazon](#). EMR
- O recurso Amazon EMR Notebooks usado com a representação de usuário do Livy não funciona porque o HTTPFs está desativado por padrão. Nesse caso, o EMR notebook não pode se conectar ao cluster que tem a representação Livy ativada. A solução alternativa é iniciar o servidor HTTPFs antes de conectar o EMR notebook ao cluster usando o `sudo systemctl start hadoop-httpfs`
- Na EMR versão 6.4.0 da Amazon, o Phoenix não oferece suporte ao componente de conectores Phoenix.
- Para usar as ações do Spark com o Apache Oozie, você deve adicionar a seguinte configuração ao seu arquivo `workflow.xml` do Oozie. Caso contrário, várias bibliotecas críticas, como a Hadoop e a Hadoop, EMRFS estarão ausentes do classpath dos executores do Spark que o Oozie lança.

```
<spark-opts>--conf spark.yarn.populateHadoopClasspath=true</spark-opts>
```

- Quando você usa o Spark com a formatação de localização de partições do Hive para ler dados no Amazon S3 e executa o Spark nas EMR versões 5.30.0 a 5.36.0 e 6.2.0 a 6.9.0 da Amazon, você pode encontrar um problema que impede seu cluster de ler os dados corretamente. Isso poderá acontecer se suas partições tiverem todas as características a seguir:
 - Duas ou mais partições são verificadas na mesma tabela.
 - Pelo menos um caminho de diretório de partição é um prefixo de pelo menos outro caminho de diretório de partição, por exemplo, `s3://bucket/table/p=a` é um prefixo de `s3://bucket/table/p=a b`.

- O primeiro caractere que segue o prefixo no outro diretório de partição tem um valor UTF -8 menor que o / caractere (U+002F). Por exemplo, o caractere de espaço (U+0020) que ocorre entre a e b em `s3://bucket/table/p=a b` se enquadra nessa categoria. Observe que existem 14 outros caracteres que não são de controle: `!"#$%&'()*+,-.` Para obter mais informações, consulte a [tabela de codificação UTF -8 e caracteres Unicode](#).

Como solução alternativa para esse problema, defina a configuração `spark.sql.sources.fastS3PartitionDiscovery.enabled` como `false` na classificação `spark-defaults`.

Versão 5.32.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.32.0 da Amazon. As alterações são referentes à versão 5.31.0.

Data da versão inicial: 8 de janeiro de 2021

Atualizações

- Atualizado o conector do Amazon Glue para a versão 1.14.0
- Amazon SageMaker Spark atualizado SDK para a versão 1.4.1
- Atualizado para AWS SDK for Java a versão 1.11.890
- Atualização do EMR DynamoDB Connector versão 4.16.0
- Atualizado EMRFS para a versão 2.45.0
- Métricas do EMR Log Analytics atualizadas para a versão 1.18.0
- EMR MetricsAndEventsApiGateway Cliente atualizado para a versão 1.5.0
- Servidor de EMR registros atualizado para a versão 1.8.0
- EMRS3 Dist CP atualizado para a versão 2.17.0
- Agente EMR secreto atualizado para a versão 1.7.0
- Atualizado o Flink para a versão 1.11.2
- Atualizado o Hadoop para a versão 2.10.1-amzn-0
- Atualizado o Hive para a versão 2.3.7-amzn-3
- Atualizado o Hue para a versão 4.8.0

- Atualizado o Mxnet para a versão 1.7.0
- Atualizado o OpenCV para a versão 4.4.0
- Atualizado o Presto para a versão 0.240.1-amzn-0
- Atualizado o Spark para a versão 2.4.7-amzn-0
- Atualizado TensorFlow para a versão 2.3.1

Alterações, melhorias e problemas resolvidos

- Esta é uma versão para corrigir problemas com o Amazon EMR Scaling quando ele falha na escalabilidade ou redução de escala de um cluster com sucesso ou causa falhas no aplicativo.
- Corrigido um problema em que as solicitações de escalabilidade falhavam em um cluster grande e altamente utilizado quando os daemons no cluster da EMR Amazon estavam executando atividades de verificação de integridade, como coletar o estado do nó e o estado do YARN nó. HDFS Isso estava acontecendo porque os daemons no cluster não conseguiam comunicar os dados do status de saúde de um nó aos componentes internos da Amazon. EMR
- Daemons EMR no cluster aprimorados para rastrear corretamente os estados dos nós quando os endereços IP são reutilizados para melhorar a confiabilidade durante as operações de escalabilidade.
- [SPARK-29683](#). Corrigido um problema em que ocorriam falhas de trabalho durante a redução da escala verticalmente do cluster, pois o Spark presumia que todos os nós disponíveis estavam na lista de negação.
- [YARN-9011](#). Corrigido um problema em que falhas de trabalho ocorriam devido a uma condição de corrida no YARN descomissionamento quando o cluster tentava aumentar ou diminuir a escala.
- Foi corrigido o problema com falhas de etapas ou tarefas durante a escalabilidade do cluster, garantindo que os estados dos nós fossem sempre consistentes entre os EMR daemons no cluster da Amazon e/. YARN HDFS
- Corrigido um problema em que operações de cluster, como redução de escala e envio de etapas, falhavam em EMR clusters da Amazon habilitados com a autenticação Kerberos. Isso ocorreu porque o daemon no EMR cluster da Amazon não renovou o tíquete Kerberos, que é necessário para se comunicar com segurança com/em execução no nó primário. HDFS YARN
- As EMR versões mais recentes da Amazon corrigem o problema com um limite menor de “Máximo de arquivos abertos” em relação aos mais antigos AL2 na AmazonEMR. As EMR versões 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 e posteriores agora incluem uma correção permanente com uma configuração mais alta de “Máximo de arquivos abertos”.

- Atualizadas versões do componente.
- Para obter uma lista das versões dos componentes, consulte [Sobre os EMR lançamentos da Amazon](#) neste guia.

Novos atributos

- A partir do Amazon EMR 5.32.0 e 6.5.0, o dimensionamento dinâmico do executor para o Apache Spark está habilitado por padrão. Para ativar ou desativar esse atributo, você pode usar o parâmetro de configuração `spark.yarn.heterogeneousExecutors.enabled`.
- Status de suporte do Instance Metadata Service (IMDS) V2: os componentes Amazon EMR 5.23.1, 5.27.1 e 5.32 ou posteriores são usados para todas as chamadas. IMDSv2 IMDS Para IMDS chamadas no código do seu aplicativo, você pode usar ambos IMDSv1 e IMDSv2, ou configurar o IMDS para uso somente IMDSv2 para aumentar a segurança. Para outras EMR versões 5.x, a desativação IMDSv1 causa falha na inicialização do cluster.
- A partir do Amazon EMR 5.32.0, você pode iniciar um cluster que se integra nativamente com o Apache Ranger. O Apache Ranger é uma estrutura de código aberto para habilitar, monitorar e gerenciar uma segurança de dados abrangente em toda a plataforma Hadoop. Para obter mais informações, consulte [Apache Ranger](#). Com a integração nativa, você pode trazer seu próprio Apache Ranger para impor um controle refinado de acesso aos dados na Amazon. EMR Consulte [Integrar a Amazon EMR com o Apache Ranger](#) no Guia de EMR lançamento da Amazon.
- A EMR versão 5.32.0 da Amazon oferece suporte à Amazon EMR on. EKS Para obter mais detalhes sobre como começar a usar o EMR onEKS, consulte [What is Amazon EMR on EKS](#).
- A EMR versão 5.32.0 da Amazon é compatível com o Amazon EMR Studio (versão prévia). Para obter mais detalhes sobre como começar a usar o EMR Studio, consulte [Amazon EMR Studio \(versão prévia\)](#).
- Políticas gerenciadas com escopo definido: para se alinhar às AWS melhores práticas, a Amazon EMR introduziu políticas EMR gerenciadas padrão com escopo v2 como substitutas das políticas que serão descontinuadas. Consulte as [políticas EMR gerenciadas da Amazon](#).

Problemas conhecidos

- Para clusters de sub-rede privados Amazon EMR 6.3.0 e 6.2.0, você não pode acessar a interface web do Ganglia. Você receberá um erro de “acesso negado (403)”. Outros sites UIs, como Spark, Hue, Zeppelin JupyterHub, Livy e Tez, estão funcionando normalmente. O acesso à interface do usuário da Web do Ganglia em clusters de sub-redes públicas também está funcionando

normalmente. Para resolver esse problema, reinicie o serviço `httpd` no nó primário com `sudo systemctl restart httpd`. Esse problema foi corrigido no Amazon EMR 6.4.0.

- Limite inferior de “Máximo de arquivos abertos” em versões mais antigas AL2 [corrigido em versões mais recentes]. EMRLançamentos da Amazon: `emr-5.30.x`, `emr-5.31.0`, `emr-5.32.0`, `emr-6.0.0`, `emr-6.1.0` e `emr-6.2.0` são baseados em versões mais antigas do Linux of Amazon 2 (AL2), que têm uma configuração de limite inferior para “Máximo de arquivos abertos” quando os clusters da Amazon são criados com o padrão. EMR AMI As EMR versões 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 e posteriores incluem uma correção permanente com uma configuração maior de “Máximo de arquivos abertos”. Versões com o limite inferior de arquivos abertos causam o erro “Muitos arquivos abertos” ao ser enviado um trabalho do Spark. Nas versões afetadas, o padrão da Amazon AMI tem uma configuração EMR padrão de `ulimit` de 4096 para “Máximo de arquivos abertos”, que é inferior ao limite de 65536 arquivos no Linux 2. latest Amazon AMI A configuração inferior de `ulimit` para “Máximo de arquivos abertos” causa falhas em trabalhos do Spark quando o driver e o executor do Spark tentam abrir mais de 4096 arquivos. Para corrigir o problema, a Amazon EMR tem um script de ação de bootstrap (BA) que ajusta a configuração `ulimit` na criação do cluster.

Se você estiver usando uma EMR versão mais antiga da Amazon que não tem a solução permanente para esse problema, a solução alternativa a seguir permite que você defina explicitamente o `ulimit` do controlador de instância para um máximo de 65536 arquivos.

Defina explicitamente um `ulimit` na linha de comando

1. Edite `/etc/systemd/system/instance-controller.service` para adicionar os seguintes parâmetros à seção Serviço.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Reiniciar InstanceController


```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Defina um `ulimit` usando a ação de bootstrap (BA)

Você também pode usar um script de ação de bootstrap (BA) para configurar o ulimit instance-controller para 65536 arquivos na criação do 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

EMRclusters que executam Amazon Linux ou Amazon Linux 2 Amazon Machine Images (AMIs) usam o comportamento padrão do Amazon Linux e não baixam e instalam automaticamente atualizações importantes e críticas do kernel que exigem uma reinicialização. Esse é o mesmo comportamento de outras EC2 instâncias da Amazon que executam o Amazon Linux padrãoAMI. Se novas atualizações de software Amazon Linux que exigem uma reinicialização (como kernel e CUDA atualizações) ficarem disponíveis após a disponibilização de uma EMR versão da Amazon, as instâncias de EMR cluster que executam o padrão AMI não baixam e instalam automaticamente essas atualizações. NVIDIA Para obter atualizações do kernel, você pode [personalizar sua Amazon EMR AMI](#) para [usar o Amazon Linux AMI mais recente](#).

- O suporte do console para criar uma configuração de segurança que especifica a opção de integração do AWS Ranger atualmente não é suportado na GovCloud região. A configuração de segurança pode ser feita usando CLI o. Consulte [Criar a configuração EMR de segurança](#) no Guia EMR de gerenciamento da Amazon.

- Quando AtRestEncryption a HDFS criptografia é habilitada em um cluster que usa o Amazon EMR 5.31.0 ou 5.32.0, as consultas do Hive resultam na seguinte exceção de tempo de execução.

```
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
```

- Quando você usa o Spark com a formatação de localização de partições do Hive para ler dados no Amazon S3 e executa o Spark nas EMR versões 5.30.0 a 5.36.0 e 6.2.0 a 6.9.0 da Amazon, você pode encontrar um problema que impede seu cluster de ler os dados corretamente. Isso poderá acontecer se suas partições tiverem todas as características a seguir:
 - Duas ou mais partições são verificadas na mesma tabela.
 - Pelo menos um caminho de diretório de partição é um prefixo de pelo menos outro caminho de diretório de partição, por exemplo, s3://bucket/table/p=a é um prefixo de s3://bucket/table/p=a b.
 - O primeiro caractere que segue o prefixo no outro diretório de partição tem um valor UTF -8 menor que o / caractere (U+002F). Por exemplo, o caractere de espaço (U+0020) que ocorre entre a e b em s3://bucket/table/p=a b se enquadra nessa categoria. Observe que existem 14 outros caracteres que não são de controle: !"#\$%&'()*+,-. Para obter mais informações, consulte a [tabela de codificação UTF -8 e caracteres Unicode](#).

Como solução alternativa para esse problema, defina a configuração `spark.sql.sources.fastS3PartitionDiscovery.enabled` como `false` na classificação `spark-defaults`.

Versão 6.2.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 6.2.0 da Amazon. As alterações são referentes à versão 6.1.0.

Data da versão inicial: 9 de dezembro de 2020

Data da última atualização: 4 de outubro de 2021

Aplicações compatíveis

- AWS SDK for Java versão 1.11.828
- emr-record-server versão 1.7.0
- Flink versão 1.11.2
- Ganglia versão 3.7.2
- Hadoop versão 3.2.1-amzn-1
- HBase versão 2.2.6-amzn-0
- HBase-operator-tools 1.0.0
- HCatalog versão 3.1.2-amzn-0
- Hive versão 3.1.2-amzn-3
- Hudi versão 0.6.0-amzn-1
- Hue versão 4.8.0
- JupyterHub versão 1.1.0
- Livy versão 0.7.0
- MXNet versão 1.7.0
- Oozie versão 5.2.0
- Phoenix versão 5.0.0
- Pig versão 0.17.0
- Presto versão 0.238.3-amzn-1
- Presto SQL versão 343
- Spark versão 3.0.1-amzn-0
- spark-rapids 0.2.0
- TensorFlow versão 2.3.1
- Zeppelin versão 0.9.0-preview1
- Zookeeper versão 3.4.14
- Conectores e drivers: DynamoDB Connector 4.16.0

Novos atributos

- HBase: a renomeação foi removida na fase de confirmação e o HFile rastreamento persistente foi adicionado. Consulte [HFileRastreamento persistente](#) no Guia EMR de lançamento da Amazon.

- HBase: Backported [Crie uma configuração que força o armazenamento em cache dos blocos na compactação](#).
- PrestoDB: melhorias na remoção dinâmica de partições. O Join Reorder baseado em regras funciona em dados não particionados.
- Políticas gerenciadas com escopo definido: para se alinhar às AWS melhores práticas, a Amazon EMR introduziu políticas EMR gerenciadas padrão com escopo v2 como substitutas das políticas que serão descontinuadas. Consulte as [políticas EMR gerenciadas da Amazon](#).
- Status de suporte do Instance Metadata Service (IMDS) V2: Para o Amazon EMR 6.2 ou posterior, os EMR componentes da Amazon são usados IMDSv2 para todas as chamadas. IMDS Para IMDS chamadas no código do seu aplicativo, você pode usar ambos IMDSv1 e IMDSv2, ou configurar o IMDS para uso somente IMDSv2 para aumentar a segurança. Se você desabilitar IMDSv1 em versões anteriores do Amazon EMR 6.x, isso causará uma falha na inicialização do cluster.

Alterações, melhorias e problemas resolvidos

- Esta é uma versão para corrigir problemas com o Amazon EMR Scaling quando ele falha na escalabilidade ou redução de escala de um cluster com sucesso ou causa falhas no aplicativo.
- Corrigido um problema em que as solicitações de escalabilidade falhavam em um cluster grande e altamente utilizado quando os daemons no cluster da EMR Amazon estavam executando atividades de verificação de integridade, como coletar o estado do nó e o estado do YARN nó. HDFS Isso estava acontecendo porque os daemons no cluster não conseguiam comunicar os dados do status de saúde de um nó aos componentes internos da Amazon. EMR
- Daemons EMR no cluster aprimorados para rastrear corretamente os estados dos nós quando os endereços IP são reutilizados para melhorar a confiabilidade durante as operações de escalabilidade.
- [SPARK-29683](#). Corrigido um problema em que ocorriam falhas de trabalho durante a redução da escala verticalmente do cluster, pois o Spark presumia que todos os nós disponíveis estavam na lista de negação.
- [YARN-9011](#). Corrigido um problema em que falhas de trabalho ocorriam devido a uma condição de corrida no YARN descomissionamento quando o cluster tentava aumentar ou diminuir a escala.
- Foi corrigido o problema com falhas de etapas ou tarefas durante a escalabilidade do cluster, garantindo que os estados dos nós fossem sempre consistentes entre os EMR daemons no cluster da Amazon e/. YARN HDFS
- Corrigido um problema em que operações de cluster, como redução de escala e envio de etapas, falhavam em EMR clusters da Amazon habilitados com a autenticação Kerberos. Isso ocorreu

porque o daemon no EMR cluster da Amazon não renovou o tíquete Kerberos, que é necessário para se comunicar com segurança com/em execução no nó primário. HDFS YARN

- As EMR versões mais recentes da Amazon corrigem o problema com um limite menor de “Máximo de arquivos abertos” em relação aos mais antigos AL2 na AmazonEMR. As EMR versões 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 e posteriores agora incluem uma correção permanente com uma configuração mais alta de “Máximo de arquivos abertos”.
- Spark: melhorias de performance no runtime do Spark.

Problemas conhecidos

- O Amazon EMR 6.2 tem permissões incorretas definidas no arquivo `/etc/cron.d/libinstance-controller-java` na versão 6.2.0. EMR As permissões no arquivo são 645 (-rw-r--r-x), quando deveriam ser 644 (-rw-r--r-). Como resultado, a EMR versão 6.2 da Amazon não registra registros do estado da instância e o diretório `/emr/instance-logs` está vazio. Esse problema foi corrigido no Amazon EMR 6.3.0 e versões posteriores.

Para contornar esse problema, execute o script a seguir como uma ação de bootstrap na inicialização do cluster.

```
#!/bin/bash
sudo chmod 644 /etc/cron.d/libinstance-controller-java
```

- Para clusters de sub-rede privados Amazon EMR 6.2.0 e 6.3.0, você não pode acessar a interface web do Ganglia. Você receberá um erro de “acesso negado (403)”. Outros sitesUIs, como Spark, Hue, Zeppelin JupyterHub, Livy e Tez, estão funcionando normalmente. O acesso à interface do usuário da Web do Ganglia em clusters de sub-redes públicas também está funcionando normalmente. Para resolver esse problema, reinicie o serviço `httpd` no nó primário com `sudo systemctl restart httpd`. Esse problema foi corrigido no Amazon EMR 6.4.0.
- Há um problema no Amazon EMR 6.2.0 em que o `httpd` falha continuamente, fazendo com que o Ganglia fique indisponível. Você recebe a mensagem de erro “cannot connect to the server”. Para corrigir um cluster que já está em execução com esse problema, SSH acesse o nó primário do cluster e adicione a linha `Listen 80` ao arquivo `httpd.conf` localizado em `/etc/httpd/conf/httpd.conf`. Esse problema foi corrigido no Amazon EMR 6.3.0.
- HTTPDfalha nos clusters EMR 6.2.0 quando você usa uma configuração de segurança. Isso faz com que a interface de usuário da aplicação Web do Ganglia fique indisponível. Para acessar a interface de usuário da aplicação Web do Ganglia, adicione `Listen 80` ao arquivo `/etc/httpd/`

conf/httpd.conf no nó primário do cluster. Para obter informações sobre como se conectar ao seu cluster, consulte [Conectar-se ao nó primário usando SSH](#).

EMROs notebooks também não conseguem estabelecer uma conexão com clusters EMR 6.2.0 quando você usa uma configuração de segurança. O caderno não conseguirá listar os kernels e enviar trabalhos do Spark. Em EMR vez disso, recomendamos que você use EMR Notebooks com outra versão da Amazon.

- Limite inferior de “Máximo de arquivos abertos” em versões mais antigas AL2 [corrigido em versões mais recentes]. EMRLançamentos da Amazon: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 e emr-6.2.0 são baseados em versões mais antigas do Linux ofAmazon 2 ()AL2, que têm uma configuração de limite inferior para “Máximo de arquivos abertos” quando os clusters da Amazon são criados com o padrão. EMR AMI As EMR versões 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 e posteriores incluem uma correção permanente com uma configuração maior de “Máximo de arquivos abertos”. Versões com o limite inferior de arquivos abertos causam o erro “Muitos arquivos abertos” ao ser enviado um trabalho do Spark. Nas versões afetadas, o padrão da Amazon AMI tem uma configuração EMR padrão de ulimit de 4096 para “Máximo de arquivos abertos”, que é inferior ao limite de 65536 arquivos no Linux 2. latestAmazon AMI A configuração inferior de ulimit para “Máximo de arquivos abertos” causa falhas em trabalhos do Spark quando o driver e o executor do Spark tentam abrir mais de 4096 arquivos. Para corrigir o problema, a Amazon EMR tem um script de ação de bootstrap (BA) que ajusta a configuração ulimit na criação do cluster.

Se você estiver usando uma EMR versão mais antiga da Amazon que não tem a solução permanente para esse problema, a solução alternativa a seguir permite que você defina explicitamente o ulimit do controlador de instância para um máximo de 65536 arquivos.

Defina explicitamente um ulimit na linha de comando

1. Edite `/etc/systemd/system/instance-controller.service` para adicionar os seguintes parâmetros à seção Serviço.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Reiniciar InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Defina um ulimit usando a ação de bootstrap (BA)

Você também pode usar um script de ação de bootstrap (BA) para configurar o ulimit instance-controller para 65536 arquivos na criação do 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

O Amazon EMR 6.1.0 e 6.2.0 incluem um problema de desempenho que pode afetar criticamente todas as operações de inserção, atualização e exclusão do Hudi. Se você planeja usar o Hudi com o Amazon EMR 6.1.0 ou 6.2.0, entre em contato com o AWS suporte para obter um Hudi corrigido. RPM

Important

EMRclusters que executam Amazon Linux ou Amazon Linux 2 Amazon Machine Images (AMIs) usam o comportamento padrão do Amazon Linux e não baixam e instalam automaticamente atualizações importantes e críticas do kernel que exigem uma reinicialização. Esse é o mesmo comportamento de outras EC2 instâncias da Amazon que executam o Amazon Linux padrãoAMI. Se novas atualizações de software Amazon Linux que exigem uma reinicialização (como kernel e CUDA atualizações) ficarem disponíveis

após a disponibilização de uma EMR versão da Amazon, as instâncias de EMR cluster que executam o padrão AMI não baixam e instalam automaticamente essas atualizações. NVIDIA Para obter atualizações do kernel, você pode [personalizar sua Amazon EMR AMI](#) para [usar o Amazon Linux AMI mais recente](#).

- Os artefatos do Amazon EMR 6.2.0 Maven não são publicados. Eles serão publicados com uma versão futura da AmazonEMR.
- O HFile rastreamento persistente usando a tabela do sistema HBase storefile não oferece suporte ao recurso de replicação da HBase região. Para obter mais informações sobre a replicação HBase da região, consulte Leituras de alta disponibilidade [consistentes com a linha do tempo](#).
- Diferenças entre as versões de bucketing do Amazon EMR EMR 6.x e 5.x Hive

EMR 5.x usa o OOS Apache Hive 2, enquanto no EMR 6.x usa OOS o Apache Hive 3. O Hive2 de código aberto usa o Bucketing versão 1, enquanto o Hive3 de código aberto usa o Bucketing versão 2. Essa diferença de versão de compartimentação entre o Hive 2 (EMR5.x) e o Hive 3 (EMR6.x) significa que o hashing de compartimentação do Hive funciona de forma diferente. Veja o exemplo abaixo.

A tabela a seguir é um exemplo criado em EMR 6.x e EMR 5.x, respectivamente.

```
-- 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/';
```

Inserindo os mesmos dados em EMR 6.x e 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');
```

Verificar a localização do S3 mostra que o nome do arquivo de compartimento é diferente, porque a função de hash é diferente entre EMR 6.x (Hive 3) e 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

```

Você também pode ver a diferença de versão executando o seguinte comando no Hive CLI na versão EMR 6.x. Observe que ele retorna a versão 2 do bucketing.

```

hive> DESCRIBE FORMATTED test_bucketing;
...
Table Parameters:
  bucketing_version      2
...

```

- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSH como hadoop usuário do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do hadoop.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo keytab está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas

definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

- Quando você usa o Spark com a formatação de localização de partições do Hive para ler dados no Amazon S3 e executa o Spark nas EMR versões 5.30.0 a 5.36.0 e 6.2.0 a 6.9.0 da Amazon, você pode encontrar um problema que impede seu cluster de ler os dados corretamente. Isso poderá acontecer se suas partições tiverem todas as características a seguir:
 - Duas ou mais partições são verificadas na mesma tabela.
 - Pelo menos um caminho de diretório de partição é um prefixo de pelo menos outro caminho de diretório de partição, por exemplo, `s3://bucket/table/p=a` é um prefixo de `s3://bucket/table/p=a b`.
 - O primeiro caractere que segue o prefixo no outro diretório de partição tem um valor UTF -8 menor que o / caractere (U+002F). Por exemplo, o caractere de espaço (U+0020) que ocorre entre a e b em `s3://bucket/table/p=a b` se enquadra nessa categoria. Observe que existem 14 outros caracteres que não são de controle: `!"#$%&'()*+,-.` Para obter mais informações, consulte a [tabela de codificação UTF -8 e caracteres Unicode](#).

Como solução alternativa para esse problema, defina a configuração `spark.sql.sources.fastS3PartitionDiscovery.enabled` como `false` na classificação `spark-defaults`.

Versão 5.31.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.31.0 da Amazon. As alterações são referentes à versão 5.30.1.

Data da versão inicial: 9 de outubro de 2020

Data da última atualização: 15 de outubro de 2020

Atualizações

- Atualizado o conector do Amazon Glue para a versão 1.13.0
- Amazon SageMaker Spark atualizado SDK para a versão 1.4.0
- Atualizado o conector do Amazon Kinesis para a versão 3.5.9
- Atualizado para AWS SDK for Java a versão 1.11.852
- Atualizado o Bigtop-tomcat para a versão 8.5.56

- EMRFS atualizado para a versão 2.43.0
- EMR MetricsAndEventsApiGateway Cliente atualizado para a versão 1.4.0
- EMRS3 Dist CP atualizado para a versão 2.15.0
- EMRS3 Select atualizado para a versão 1.6.0
- Atualizado o Flink para a versão 1.11.0
- Atualizado o Hadoop para a versão 2.10.0
- Atualizado o Hive para a versão 2.3.7
- Atualizado o Hudi para a versão 0.6.0
- Atualizado o Hue para a versão 4.7.1
- Atualizado JupyterHub para a versão 1.1.0
- Atualizado o Mxnet para a versão 1.6.0
- Atualizado o OpenCV para a versão 4.3.0
- Atualizado o Presto para a versão 0.238.3
- Atualizado TensorFlow para a versão 2.1.0

Alterações, melhorias e problemas resolvidos

- Esta é uma versão para corrigir problemas com o Amazon EMR Scaling quando ele falha na escalabilidade ou redução de escala de um cluster com sucesso ou causa falhas no aplicativo.
- Corrigido um problema em que as solicitações de escalabilidade falhavam em um cluster grande e altamente utilizado quando os daemons no cluster da EMR Amazon estavam executando atividades de verificação de integridade, como coletar o estado do nó e o estado do YARN nó. HDFS Isso estava acontecendo porque os daemons no cluster não conseguiam comunicar os dados do status de saúde de um nó aos componentes internos da Amazon. EMR
- Daemons EMR no cluster aprimorados para rastrear corretamente os estados dos nós quando os endereços IP são reutilizados para melhorar a confiabilidade durante as operações de escalabilidade.
- [SPARK-29683](#). Corrigido um problema em que ocorriam falhas de trabalho durante a redução da escala verticalmente do cluster, pois o Spark presumia que todos os nós disponíveis estavam na lista de negação.
- [YARN-9011](#). Corrigido um problema em que falhas de trabalho ocorriam devido a uma condição de corrida no YARN descomissionamento quando o cluster tentava aumentar ou diminuir a escala.

- Foi corrigido o problema com falhas de etapas ou tarefas durante a escalabilidade do cluster, garantindo que os estados dos nós fossem sempre consistentes entre os EMR daemons no cluster da Amazon e/. YARN HDFS
- Corrigido um problema em que operações de cluster, como redução de escala e envio de etapas, falhavam em EMR clusters da Amazon habilitados com a autenticação Kerberos. Isso ocorreu porque o daemon no EMR cluster da Amazon não renovou o tíquete Kerberos, que é necessário para se comunicar com segurança com/em execução no nó primário. HDFS YARN
- As EMR versões mais recentes da Amazon corrigem o problema com um limite menor de “Máximo de arquivos abertos” em relação aos mais antigos AL2 na AmazonEMR. As EMR versões 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 e posteriores agora incluem uma correção permanente com uma configuração mais alta de “Máximo de arquivos abertos”.
- [As estatísticas de colunas do Hive](#) são compatíveis com as EMR versões 5.31.0 e posteriores da Amazon.
- Atualizadas versões do componente.
- EMRFSSuporte ao S3EC V2 na Amazon EMR 5.31.0. Nas SDK versões 1.11.837 e posteriores do Java do S3, a versão 2 do cliente de criptografia (S3EC V2) foi introduzida com vários aprimoramentos de segurança. Para obter mais informações, consulte as informações a seguir.
 - Publicações no blog do S3: [Atualizações no cliente de criptografia do Amazon S3](#).
 - AWS SDK for Java Guia do desenvolvedor: [Migre clientes de criptografia e descriptografia](#) para a V2.
 - EMRGuia de gerenciamento: criptografia [do lado do cliente do Amazon S3](#).

O Encryption Client V1 ainda está disponível no SDK para compatibilidade com versões anteriores.

Novos atributos

- Limite inferior de “Máximo de arquivos abertos” em versões mais antigas AL2 [corrigido em versões mais recentes]. EMRLançamentos da Amazon: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 e emr-6.2.0 são baseados em versões mais antigas do Linux ofAmazon 2 ()AL2, que têm uma configuração de limite inferior para “Máximo de arquivos abertos” quando os clusters da Amazon são criados com o padrão. EMR AMI As EMR versões 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 e posteriores incluem uma correção permanente com uma configuração maior de “Máximo de arquivos abertos”. Versões com o limite inferior de arquivos abertos causam o erro “Muitos arquivos abertos” ao ser enviado um trabalho do Spark. Nas versões afetadas, o padrão da Amazon AMI tem uma configuração EMR padrão de ulimit

de 4096 para “Máximo de arquivos abertos”, que é inferior ao limite de 65536 arquivos no Linux 2. latestAmazon AMI A configuração inferior de ulimit para “Máximo de arquivos abertos” causa falhas em trabalhos do Spark quando o driver e o executor do Spark tentam abrir mais de 4096 arquivos. Para corrigir o problema, a Amazon EMR tem um script de ação de bootstrap (BA) que ajusta a configuração ulimit na criação do cluster.

Se você estiver usando uma EMR versão mais antiga da Amazon que não tem a solução permanente para esse problema, a solução alternativa a seguir permite que você defina explicitamente o ulimit do controlador de instância para um máximo de 65536 arquivos.

Defina explicitamente um ulimit na linha de comando

1. Edite `/etc/systemd/system/instance-controller.service` para adicionar os seguintes parâmetros à seção Serviço.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Reiniciar InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Defina um ulimit usando a ação de bootstrap (BA)

Você também pode usar um script de ação de bootstrap (BA) para configurar o ulimit instance-controller para 65536 arquivos na criação do 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
```

- Com o Amazon EMR 5.31.0, você pode lançar um cluster que se integra ao Lake Formation. Essa integração fornece filtragem de dados refinada em nível de coluna para bancos de dados e tabelas no Glue Data Catalog. AWS Ele também permite o login único federado em EMR notebooks ou Apache Zeppelin a partir de um sistema de identidade corporativo. Para obter mais informações, consulte [Integração da Amazon EMR com](#) o Guia AWS Lake Formation de EMR gerenciamento da Amazon.

Atualmente, o Amazon EMR with Lake Formation está disponível em 16 AWS regiões: Leste dos EUA (Ohio e Norte da Virgínia), Oeste dos EUA (Norte da Califórnia e Oregon), Ásia-Pacífico (Mumbai, Seul, Cingapura, Sydney e Tóquio), Canadá (Central), Europa (Frankfurt, Irlanda, Londres, Paris e Estocolmo) e América do Sul (São Paulo).

Problemas conhecidos

- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSH como hadoop usuário do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do hadoop.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo `keytab` está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

- Quando `AtRestEncryption` a HDFS criptografia é habilitada em um cluster que usa o Amazon EMR 5.31.0 ou 5.32.0, as consultas do Hive resultam na seguinte exceção de tempo de execução.

```
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
```

- Quando você usa o Spark com a formatação de localização de partições do Hive para ler dados no Amazon S3 e executa o Spark nas EMR versões 5.30.0 a 5.36.0 e 6.2.0 a 6.9.0 da Amazon, você pode encontrar um problema que impede seu cluster de ler os dados corretamente. Isso poderá acontecer se suas partições tiverem todas as características a seguir:
 - Duas ou mais partições são verificadas na mesma tabela.
 - Pelo menos um caminho de diretório de partição é um prefixo de pelo menos outro caminho de diretório de partição, por exemplo, `s3://bucket/table/p=a` é um prefixo de `s3://bucket/table/p=a b`.
 - O primeiro caractere que segue o prefixo no outro diretório de partição tem um valor UTF -8 menor que o `/` caractere (U+002F). Por exemplo, o caractere de espaço (U+0020) que ocorre entre `a` e `b` em `s3://bucket/table/p=a b` se enquadra nessa categoria. Observe que existem 14 outros caracteres que não são de controle: `!"#$%&'()*+,-.` Para obter mais informações, consulte a [tabela de codificação UTF -8 e caracteres Unicode](#).

Como solução alternativa para esse problema, defina a configuração `spark.sql.sources.fastS3PartitionDiscovery.enabled` como `false` na classificação `spark-defaults`.

Versão 6.1.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 6.1.0 da Amazon. As alterações são referentes à versão 6.0.0.

Data da versão inicial: 4 de setembro de 2020

Data da última atualização: 15 de outubro de 2020

Aplicações compatíveis

- AWS SDK for Java versão 1.11.828
- Flink versão 1.11.0
- Ganglia versão 3.7.2
- Hadoop versão 3.2.1-amzn-1
- HBase versão 2.2.5
- HBase-operator-tools 1.0.0
- HCatalog versão 3.1.2-amzn-0
- Hive versão 3.1.2-amzn-1
- Hudi versão 0.5.2-incubating
- Hue versão 4.7.1
- JupyterHub versão 1.1.0
- Livy versão 0.7.0
- MXNet versão 1.6.0
- Oozie versão 5.2.0
- Phoenix versão 5.0.0
- Presto versão 0.232
- Presto SQL versão 3.38
- Spark versão 3.0.0-amzn-0
- TensorFlow versão 2.1.0
- Zeppelin versão 0.9.0-preview1
- Zookeeper versão 3.4.14

- Conectores e drivers: DynamoDB Connector 4.14.0

Novos atributos

- ARMOs tipos de instância são suportados a partir da Amazon EMR versão 5.30.0 e da Amazon EMR versão 6.1.0.
- Os tipos de instância de uso geral M6g são suportados a partir das EMR versões 6.1.0 e 5.30.0 da Amazon. Para obter mais informações, consulte [Tipos de instância compatíveis](#) no Amazon EMR Management Guide.
- O recurso de grupos de EC2 posicionamento é suportado a partir da EMR versão 5.23.0 da Amazon como uma opção para vários clusters de nós primários. Atualmente, somente os tipos de nós primários são compatíveis com o atributo grupo de posicionamento e a estratégia SPREAD é aplicada a estes nós primários. A estratégia SPREAD posiciona um pequeno grupo de instâncias em um hardware subjacente separado para evitar a perda de múltiplos nós primários em caso de falha de hardware. Para obter mais informações, consulte [EMRIntegração com o EC2 Placement Group](#) no Amazon EMR Management Guide.
- Escalabilidade gerenciada — Com a EMR versão 6.1.0 da Amazon, você pode permitir que a escalabilidade EMR gerenciada da Amazon aumente ou diminua automaticamente o número de instâncias ou unidades em seu cluster com base na carga de trabalho. A Amazon avalia EMR continuamente as métricas do cluster para tomar decisões de escalabilidade que otimizem seus clusters em termos de custo e velocidade. O Managed Scaling também está disponível na Amazon EMR versão 5.30.0 e posterior, exceto 6.0.0. Para obter mais informações, consulte [Scaling Cluster Resources](#) no Amazon EMR Management Guide.
- A SQL versão 338 do Presto é compatível com EMR 6.1.0. Para obter mais informações, consulte [Presto](#).
 - O Presto SQL é suportado somente na EMR versão 6.1.0 e versões posteriores, não na EMR 6.0.0 ou 5.x. EMR
 - O nome da aplicação, Presto, continua a ser usado para instalar o PrestoDB em clusters. Para instalar o Presto SQL em clusters, use o nome PrestoSQL do aplicativo.
 - Você pode instalar o PrestoDB ou o SQL Presto, mas não pode instalar os dois em um único cluster. Se o PrestoDB e o SQL Presto forem especificados ao tentar criar um cluster, ocorrerá um erro de validação e a solicitação de criação do cluster falhará.
 - O Presto SQL é compatível com clusters single-master e multimaster. Em clusters com vários mestres, é necessário um metastore externo do Hive para executar o Presto ou o PrestoDB. SQL Consulte [Aplicativos compatíveis em um EMR cluster com vários nós primários](#).

- ECRsuporte à autenticação automática no Apache Hadoop e no Apache Spark com Docker: os usuários do Spark podem usar imagens do Docker Hub e do Amazon Elastic Container Registry (AmazonECR) para definir dependências de ambiente e biblioteca.

[Configure o Docker](#) e [execute aplicativos Spark com o Docker usando o Amazon 6.x](#). EMR

- EMRsuporta ACID transações do Apache Hive: o Amazon EMR 6.1.0 adiciona suporte às transações do Hive para que esteja em conformidade com ACID as propriedades de um banco de dados. ACID Com esse atributo, você pode executar as operações INSERT, UPDATE, DELETE, e MERGE em tabelas gerenciadas do Hive com dados no Amazon Simple Storage Service (Amazon S3). Esse é um recurso fundamental para casos de uso como ingestão de streaming, atualização de dados, uso de atualizações em massa e mudanças lentas de dimensões. MERGE Para obter mais informações, incluindo exemplos de configuração e casos de uso, consulte [Amazon EMR suporta transações do Apache Hive ACID](#).

Alterações, melhorias e problemas resolvidos

- Esta é uma versão para corrigir problemas com o Amazon EMR Scaling quando ele falha na escalabilidade ou redução de escala de um cluster com sucesso ou causa falhas no aplicativo.
- Corrigido um problema em que as solicitações de escalabilidade falhavam em um cluster grande e altamente utilizado quando os daemons no cluster da EMR Amazon estavam executando atividades de verificação de integridade, como coletar o estado do nó e o estado do YARN nó. HDFS Isso estava acontecendo porque os daemons no cluster não conseguiam comunicar os dados do status de saúde de um nó aos componentes internos da Amazon. EMR
- Daemons EMR no cluster aprimorados para rastrear corretamente os estados dos nós quando os endereços IP são reutilizados para melhorar a confiabilidade durante as operações de escalabilidade.
- [SPARK-29683](#). Corrigido um problema em que ocorriam falhas de trabalho durante a redução da escala verticalmente do cluster, pois o Spark presumia que todos os nós disponíveis estavam na lista de negação.
- [YARN-9011](#). Corrigido um problema em que falhas de trabalho ocorriam devido a uma condição de corrida no YARN descomissionamento quando o cluster tentava aumentar ou diminuir a escala.
- Foi corrigido o problema com falhas de etapas ou tarefas durante a escalabilidade do cluster, garantindo que os estados dos nós fossem sempre consistentes entre os EMR daemons no cluster da Amazon e/. YARN HDFS

- Corrigido um problema em que operações de cluster, como redução de escala e envio de etapas, falhavam em EMR clusters da Amazon habilitados com a autenticação Kerberos. Isso ocorreu porque o daemon no EMR cluster da Amazon não renovou o tíquete Kerberos, que é necessário para se comunicar com segurança com/em execução no nó primário. HDFS YARN
- As EMR versões mais recentes da Amazon corrigem o problema com um limite menor de “Máximo de arquivos abertos” em relação aos mais antigos AL2 na AmazonEMR. As EMR versões 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 e posteriores agora incluem uma correção permanente com uma configuração mais alta de “Máximo de arquivos abertos”.
- O Apache Flink não é suportado no EMR 6.0.0, mas é suportado no EMR 6.1.0 com o Flink 1.11.0. Esta é a primeira versão do Flink a oficialmente oferecer suporte ao Hadoop 3. Consulte o [Anúncio de versão do Apache Flink 1.11.0](#).
- O Ganglia foi removido dos pacotes padrão EMR 6.1.0.

Problemas conhecidos

- Limite inferior de “Máximo de arquivos abertos” em versões mais antigas AL2 [corrigido em versões mais recentes]. EMRLançamentos da Amazon: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 e emr-6.2.0 são baseados em versões mais antigas do Linux ofAmazon 2 (AL2, que têm uma configuração de limite inferior para “Máximo de arquivos abertos” quando os clusters da Amazon são criados com o padrão. EMR AMI As EMR versões 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 e posteriores incluem uma correção permanente com uma configuração maior de “Máximo de arquivos abertos”. Versões com o limite inferior de arquivos abertos causam o erro “Muitos arquivos abertos” ao ser enviado um trabalho do Spark. Nas versões afetadas, o padrão da Amazon AMI tem uma configuração EMR padrão de ulimit de 4096 para “Máximo de arquivos abertos”, que é inferior ao limite de 65536 arquivos no Linux 2. latestAmazon AMI A configuração inferior de ulimit para “Máximo de arquivos abertos” causa falhas em trabalhos do Spark quando o driver e o executor do Spark tentam abrir mais de 4096 arquivos. Para corrigir o problema, a Amazon EMR tem um script de ação de bootstrap (BA) que ajusta a configuração ulimit na criação do cluster.

Se você estiver usando uma EMR versão mais antiga da Amazon que não tem a solução permanente para esse problema, a solução alternativa a seguir permite que você defina explicitamente o ulimit do controlador de instância para um máximo de 65536 arquivos.

Defina explicitamente um ulimit na linha de comando

1. Edite `/etc/systemd/system/instance-controller.service` para adicionar os seguintes parâmetros à seção Serviço.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Reiniciar InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Defina um ulimit usando a ação de bootstrap (BA)

Você também pode usar um script de ação de bootstrap (BA) para configurar o ulimit instance-controller para 65536 arquivos na criação do 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**

O Amazon EMR 6.1.0 e 6.2.0 incluem um problema de desempenho que pode afetar criticamente todas as operações de inserção, atualização e exclusão do Hudi. Se você planeja usar o Hudi com o Amazon EMR 6.1.0 ou 6.2.0, entre em contato com o AWS suporte para obter um Hudi corrigido. RPM

- Se você definir uma configuração personalizada de coleta de lixo com `spark.driver.extraJavaOptions` e `spark.executor.extraJavaOptions`, isso resultará na falha de inicialização do driver/executor com a versão EMR 6.1 devido à configuração conflitante da coleta de lixo. Com a EMR versão 6.1.0, você deve especificar uma configuração personalizada de coleta de lixo do Spark para drivers e executores com as propriedades e em vez disso `spark.driver.defaultJavaOptions` e `spark.executor.defaultJavaOptions`. Leia mais em [Apache Spark Runtime Environment](#) e [Configurando a coleta de lixo do Spark](#) na Amazon EMR 6.1.0.
- Usar o Pig com o Oozie (e dentro do Hue, já que o Hue usa ações do Oozie para executar scripts do Pig) gera um erro em que uma biblioteca nativa lzo não pode ser carregada. Essa mensagem de erro é informativa e não impede a execução do Pig.
- Suporte de simultaneidade do Hudi: atualmente, o Hudi não é compatível com gravações simultâneas em uma única tabela do Hudi. Além disso, o Hudi reverte todas as alterações feitas por gravadores em andamento antes de permitir que um novo gravador seja iniciado. As gravações simultâneas podem interferir nesse mecanismo e introduzir condições de corrida, o que pode causar corrupção de dados. Você deve garantir que, como parte do seu fluxo de trabalho de processamento de dados, só exista um gravador do Hudi operando em uma tabela do Hudi em qualquer instante. O Hudi permite vários leitores simultâneos operando na mesma tabela do Hudi.
- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSH como usuário do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do hadoop.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo keytab está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

- Há um problema no Amazon EMR 6.1.0 que afeta os clusters que executam o Presto. Depois de um longo período (dias), o cluster pode gerar erros, como “su: failed to execute /bin/bash: Resource temporarily unavailable” ou “shell request failed on channel 0”. Esse problema é causado por um EMR processo interno da Amazon (InstanceController) que está gerando muitos processos leves (LWP), o que acaba fazendo com que o usuário do Hadoop exceda o limite de nproc. Isso impede que o usuário abra processos adicionais. A solução para esse problema é atualizar para EMR 6.2.0.

Versão 6.0.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 6.0.0 da Amazon.

Data da versão inicial: 10 de março de 2020

Aplicações compatíveis

- AWS SDK for Java versão 1.11.711
- Ganglia versão 3.7.2
- Hadoop versão 3.2.1
- HBase versão 2.2.3
- HCatalog versão 3.1.2

- Hive versão 3.1.2
- Hudi versão 0.5.0 incubadora
- Hue versão 4.4.0
- JupyterHub versão 1.0.0
- Livy versão 0.6.0
- MXNet versão 1.5.1
- Oozie versão 5.1.0
- Phoenix versão 5.0.0
- Presto versão 0.230
- Spark versão 2.4.4
- TensorFlow versão 1.14.0
- Zeppelin versão 0.9.0- SNAPSHOT
- Zookeeper versão 3.4.14
- Conectores e drivers: DynamoDB Connector 4.14.0

Note

Flink, Sqoop, Pig e Mahout não estão disponíveis na versão 6.0.0 da Amazon. EMR

Novos atributos

- YARN Docker Runtime Support — YARN aplicativos, como trabalhos do Spark, agora podem ser executados no contexto de um contêiner Docker. Isso permite que você defina facilmente dependências em uma imagem do Docker sem a necessidade de instalar bibliotecas personalizadas em seu cluster da AmazonEMR. Para obter mais informações, consulte [Configurar a integração do Docker](#) e [executar aplicativos Spark com o Docker usando o Amazon 6.0.0](#). EMR
- LLAP Suporte do Hive - O Hive agora suporta o modo de LLAP execução para melhorar o desempenho das consultas. Para obter mais informações, consulte [Usando o Hive. LLAP](#)

Alterações, melhorias e problemas resolvidos

- Esta é uma versão para corrigir problemas com o Amazon EMR Scaling quando ele falha na escalabilidade ou redução de escala de um cluster com sucesso ou causa falhas no aplicativo.

- Corrigido um problema em que as solicitações de escalabilidade falhavam em um cluster grande e altamente utilizado quando os daemons no cluster da EMR Amazon estavam executando atividades de verificação de integridade, como coletar o estado do nó e o estado do YARN nó. HDFS Isso estava acontecendo porque os daemons no cluster não conseguiam comunicar os dados do status de saúde de um nó aos componentes internos da Amazon. EMR
- Daemons EMR no cluster aprimorados para rastrear corretamente os estados dos nós quando os endereços IP são reutilizados para melhorar a confiabilidade durante as operações de escalabilidade.
- [SPARK-29683](#). Corrigido um problema em que ocorriam falhas de trabalho durante a redução da escala verticalmente do cluster, pois o Spark presumia que todos os nós disponíveis estavam na lista de negação.
- [YARN-9011](#). Corrigido um problema em que falhas de trabalho ocorriam devido a uma condição de corrida no YARN descomissionamento quando o cluster tentava aumentar ou diminuir a escala.
- Foi corrigido o problema com falhas de etapas ou tarefas durante a escalabilidade do cluster, garantindo que os estados dos nós fossem sempre consistentes entre os EMR daemons no cluster da Amazon e/. YARN HDFS
- Corrigido um problema em que operações de cluster, como redução de escala e envio de etapas, falhavam em EMR clusters da Amazon habilitados com a autenticação Kerberos. Isso ocorreu porque o daemon no EMR cluster da Amazon não renovou o tíquete Kerberos, que é necessário para se comunicar com segurança com/em execução no nó primário. HDFS YARN
- As EMR versões mais recentes da Amazon corrigem o problema com um limite menor de “Máximo de arquivos abertos” em relação aos mais antigos AL2 na AmazonEMR. As EMR versões 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 e posteriores agora incluem uma correção permanente com uma configuração mais alta de “Máximo de arquivos abertos”.
- Amazon Linux
 - O Amazon Linux 2 é o sistema operacional da série de lançamento EMR 6.x.
 - `systemd` usado para gerenciamento de serviços em vez do `inAmazon Linux 1 upstart` usado.
- Kit de desenvolvimento Java (JDK)
 - Corretto JDK 8 é o padrão JDK para a série de lançamento EMR 6.x.
- Scala
 - O Scala 2.12 é usado com o Apache Spark e com o Apache Livy.
- Python 3

- O Python 3 agora é a versão padrão do Python em. EMR
- YARNrótulos de nós
 - A partir da série de lançamento Amazon EMR 6.x, o recurso de rótulos de YARN nós está desativado por padrão. Os principais processos do aplicativo podem ser executados tanto nos nós core como nos nós de tarefa por padrão. Você pode ativar o recurso de rótulos de YARN nós configurando as seguintes propriedades: `yarn.node-labels.enabled` e `yarn.node-labels.am.default-node-label-expression` Para obter mais informações, consulte [Noções básicas sobre nós de tarefa, centrais e primários](#).

Problemas conhecidos

- Limite inferior de “Máximo de arquivos abertos” em versões mais antigas AL2 [corrigido em versões mais recentes]. EMRLançamentos da Amazon: `emr-5.30.x`, `emr-5.31.0`, `emr-5.32.0`, `emr-6.0.0`, `emr-6.1.0` e `emr-6.2.0` são baseados em versões mais antigas do Linux ofAmazon 2 (AL2, que têm uma configuração de limite inferior para “Máximo de arquivos abertos” quando os clusters da Amazon são criados com o padrão. EMR AMI As EMR versões `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` e posteriores incluem uma correção permanente com uma configuração maior de “Máximo de arquivos abertos”. Versões com o limite inferior de arquivos abertos causam o erro “Muitos arquivos abertos” ao ser enviado um trabalho do Spark. Nas versões afetadas, o padrão da Amazon AMI tem uma configuração EMR padrão de `ulimit` de 4096 para “Máximo de arquivos abertos”, que é inferior ao limite de 65536 arquivos no Linux 2. latestAmazon AMI A configuração inferior de `ulimit` para “Máximo de arquivos abertos” causa falhas em trabalhos do Spark quando o driver e o executor do Spark tentam abrir mais de 4096 arquivos. Para corrigir o problema, a Amazon EMR tem um script de ação de bootstrap (BA) que ajusta a configuração `ulimit` na criação do cluster.

Se você estiver usando uma EMR versão mais antiga da Amazon que não tem a solução permanente para esse problema, a solução alternativa a seguir permite que você defina explicitamente o `ulimit` do controlador de instância para um máximo de 65536 arquivos.

Defina explicitamente um `ulimit` na linha de comando

1. Edite `/etc/systemd/system/instance-controller.service` para adicionar os seguintes parâmetros à seção Serviço.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Reiniciar InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Defina um ulimit usando a ação de bootstrap (BA)

Você também pode usar um script de ação de bootstrap (BA) para configurar o ulimit instance-controller para 65536 arquivos na criação do 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
```

- O shell interativo do Spark PySpark, incluindo o SparkR e o spark-shell, não é compatível com o uso do Docker com bibliotecas adicionais.
- Para usar o Python 3 com a EMR versão 6.0.0 da Amazon, você deve adicionar a. `PATH yarn.nodemanager.env-whitelist`
- A funcionalidade Live Long and Process (LLAP) não é suportada quando você usa o AWS Glue Data Catalog como metastore do Hive.

- Ao usar o Amazon EMR 6.0.0 com a integração do Spark e do Docker, você precisa configurar as instâncias em seu cluster com o mesmo tipo de instância e a mesma quantidade de EBS volumes para evitar falhas ao enviar um trabalho do Spark com o tempo de execução do Docker.
- [No Amazon EMR 6.0.0, HBase no Amazon S3, o modo de armazenamento é afetado pelo HBASE problema -24286](#). HBaseo master não pode inicializar quando o cluster é criado usando dados existentes do S3.
- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSH como hadoop usuário do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do hadoop.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo keytab está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

Versão 5.30.1

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.30.1 da Amazon. As alterações são referentes à versão 5.30.0.

Data da versão inicial: 30 de junho de 2020

Data da última atualização: 24 de agosto de 2020

Alterações, melhorias e problemas resolvidos

- As EMR versões mais recentes da Amazon corrigem o problema com um limite menor de “Máximo de arquivos abertos” em relação aos mais antigos AL2 na AmazonEMR. As EMR versões 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 e posteriores agora incluem uma correção permanente com uma configuração mais alta de “Máximo de arquivos abertos”.
- Corrigido um problema em que o processo do controlador da instância gerava um número infinito de processos.
- Corrigido um problema em que o Hue não conseguia executar uma consulta do Hive, mostrando a mensagem “o banco de dados está bloqueado” e impedindo a execução de consultas.
- Foi corrigido um problema do Spark para permitir que mais tarefas fossem executadas simultaneamente no EMR cluster.
- Corrigido um problema no caderno Jupyter que causava um “erro de muitos arquivos abertos” no servidor Jupyter.
- Corrigido um problema com as horas de início do cluster.

Novos atributos

- As interfaces de aplicativos persistentes do Tez UI e do YARN Timeline Server estão disponíveis com EMR as versões 6.x e 5.30.1 e EMR posteriores da Amazon. O acesso por link com um clique ao histórico persistente do aplicativo permite que você acesse rapidamente o histórico de tarefas sem configurar um proxy da web por meio de uma SSH conexão. Os logs de clusters ativos e encerrados ficam disponíveis por 30 dias após o término da aplicação. Para obter mais informações, consulte [Exibir interfaces de usuário de aplicativos persistentes](#) no Amazon EMR Management Guide.
- EMRA execução do notebook APIs está disponível para executar EMR notebooks por meio de um script ou linha de comando. A capacidade de iniciar, parar, listar e descrever as execuções do EMR notebook sem o AWS console permite que você controle programaticamente um notebook. EMR Ao usar uma célula do caderno parametrizada, você pode passar valores de parâmetros diferentes para um caderno sem precisar criar uma cópia do caderno para cada novo conjunto de valores de parâmetros. Consulte [EMRAPiAções](#). Para obter um exemplo de código, consulte [Exemplos de comandos para executar EMR Notebooks programaticamente](#).

Problemas conhecidos

- Limite inferior de “Máximo de arquivos abertos” em versões mais antigas AL2 [corrigido em versões mais recentes]. EMRLançamentos da Amazon: emr-5.30.x, emr-5.31.0, emr-5.32.0, emr-6.0.0, emr-6.1.0 e emr-6.2.0 são baseados em versões mais antigas do Linux ofAmazon 2 (AL2, que têm uma configuração de limite inferior para “Máximo de arquivos abertos” quando os clusters da Amazon são criados com o padrão. EMR AMI As EMR versões 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 e posteriores incluem uma correção permanente com uma configuração maior de “Máximo de arquivos abertos”. Versões com o limite inferior de arquivos abertos causam o erro “Muitos arquivos abertos” ao ser enviado um trabalho do Spark. Nas versões afetadas, o padrão da Amazon AMI tem uma configuração EMR padrão de ulimit de 4096 para “Máximo de arquivos abertos”, que é inferior ao limite de 65536 arquivos no Linux 2. latestAmazon AMI A configuração inferior de ulimit para “Máximo de arquivos abertos” causa falhas em trabalhos do Spark quando o driver e o executor do Spark tentam abrir mais de 4096 arquivos. Para corrigir o problema, a Amazon EMR tem um script de ação de bootstrap (BA) que ajusta a configuração ulimit na criação do cluster.

Se você estiver usando uma EMR versão mais antiga da Amazon que não tem a solução permanente para esse problema, a solução alternativa a seguir permite que você defina explicitamente o ulimit do controlador de instância para um máximo de 65536 arquivos.

Defina explicitamente um ulimit na linha de comando

1. Edite `/etc/systemd/system/instance-controller.service` para adicionar os seguintes parâmetros à seção Serviço.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Reiniciar InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Defina um ulimit usando a ação de bootstrap (BA)

Você também pode usar um script de ação de bootstrap (BA) para configurar o ulimit instance-controller para 65536 arquivos na criação do 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
```

- EMRCadernos

O recurso que permite instalar kernels e bibliotecas Python adicionais no nó primário do cluster está desativado por padrão na versão 5.30.1. EMR Para obter mais informações sobre esse atributo, consulte [Instalar kernels e bibliotecas Python em um nó primário do cluster](#).

Para habilitar o recurso, faça o seguinte:

1. Certifique-se de que a política de permissões anexada à função de serviço para EMR Notebooks permita a seguinte ação:

```
elasticmapreduce:ListSteps
```

Para obter mais informações, consulte [Função de serviço para EMR notebooks](#).

2. Use o AWS CLI para executar uma etapa no cluster que configura os EMR Notebooks, conforme mostrado no exemplo a seguir. Substituir *us-east-1* com a região em que seu cluster reside. Para obter mais informações, consulte [Adding Steps to a Cluster Using the 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"]
```

- Ajuste de escala gerenciado

As operações de ajuste de escala gerenciado nos clusters das versões 5.30.0 e 5.30.1 sem o Presto instalado podem causar falhas na aplicação ou fazer com que um grupo de instâncias ou uma frota de instâncias uniforme permaneça no estado ARRESTED, sobretudo quando uma operação de redução da escala verticalmente logo é seguida por uma operação de aumento da escala verticalmente.

Como solução alternativa, escolha o Presto como um aplicativo a ser instalado ao criar um cluster com as EMR versões 5.30.0 e 5.30.1 da Amazon, mesmo que seu trabalho não exija o Presto.

- Problema conhecido em clusters com vários nós primários e autenticação Kerberos


Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSH como hadoop usuário do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do hadoop.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo keytab está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

 Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas

definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

- Quando você usa o Spark com a formatação de localização de partições do Hive para ler dados no Amazon S3 e executa o Spark nas EMR versões 5.30.0 a 5.36.0 e 6.2.0 a 6.9.0 da Amazon, você pode encontrar um problema que impede seu cluster de ler os dados corretamente. Isso poderá acontecer se suas partições tiverem todas as características a seguir:
 - Duas ou mais partições são verificadas na mesma tabela.
 - Pelo menos um caminho de diretório de partição é um prefixo de pelo menos outro caminho de diretório de partição, por exemplo, `s3://bucket/table/p=a` é um prefixo de `s3://bucket/table/p=a b`.
 - O primeiro caractere que segue o prefixo no outro diretório de partição tem um valor UTF -8 menor que o / caractere (U+002F). Por exemplo, o caractere de espaço (U+0020) que ocorre entre a e b em `s3://bucket/table/p=a b` se enquadra nessa categoria. Observe que existem 14 outros caracteres que não são de controle: !"#\$%&'()*+,-. Para obter mais informações, consulte a [tabela de codificação UTF -8 e caracteres Unicode](#).

Como solução alternativa para esse problema, defina a configuração `spark.sql.sources.fastS3PartitionDiscovery.enabled` como `false` na classificação `spark-defaults`.

Versão 5.30.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.30.0 da Amazon. As alterações são referentes à versão 5.29.0.

Data da versão inicial: 13 de maio de 2020

Data da última atualização: 25 de junho de 2020

Atualizações

- Atualizado para AWS SDK for Java a versão 1.11.759
- Amazon SageMaker Spark atualizado SDK para a versão 1.3.0
- Servidor de EMR registros atualizado para a versão 1.6.0
- Atualização do Flink para a versão 1.10.0
- Atualização do Ganglia para a versão 3.7.2

- Atualizado HBase para a versão 1.4.13
- Atualização do Hudi para a versão 0.5.2-incubating
- Atualização do Hue para a versão 4.6.0
- Atualizado JupyterHub para a versão 1.1.0
- Atualização do Livy para a versão 0.7.0-incubating
- Atualização do Oozie para a versão 5.2.0
- Atualização do Presto para a versão 0.232
- Atualização do Spark para a versão 2.4.5
- Conectores e drivers atualizados: Amazon Glue Connector 1.12.0; Amazon Kinesis Connector 3.5.0; DynamoDB Connector 4.14.0 EMR

Novos atributos

- EMRNotebooks — Quando usados com EMR clusters criados usando 5.30.0, os kernels do EMR notebook são executados no cluster. Isso melhora o desempenho do bloco de anotações e permite que instalar e personalizar kernels. Você também pode instalar bibliotecas Python no nó primário do cluster. Para obter mais informações, consulte [Instalando e usando kernels e bibliotecas](#) no Guia EMRde gerenciamento.
- Escalabilidade gerenciada — Com a Amazon EMR versão 5.30.0 e posterior, você pode habilitar a escalabilidade EMR gerenciada para aumentar ou diminuir automaticamente o número de instâncias ou unidades em seu cluster com base na carga de trabalho. A Amazon avalia EMR continuamente as métricas do cluster para tomar decisões de escalabilidade que otimizem seus clusters em termos de custo e velocidade. Para obter mais informações, consulte [Scaling Cluster Resources](#) no Amazon EMR Management Guide.
- Criptografe arquivos de log armazenados no Amazon S3 — Com a EMR Amazon versão 5.30.0 e posterior, você pode criptografar arquivos de log armazenados no Amazon S3 com uma chave gerenciada pelo cliente. AWS KMS Para obter mais informações, consulte [Criptografar arquivos de log armazenados no Amazon S3](#) no EMRAmazon Management Guide.
- Suporte ao Amazon Linux 2 — Na EMR versão 5.30.0 e posterior, EMR usesAmazon sistema operacional Linux 2. A nova personalização AMIs (Amazon Machine Image) deve ser baseada no theAmazon Linux 2AMI. Para obter mais informações, consulte [Usando um personalizado AMI](#).
- Presto Graceful Auto Scale — EMR clusters que usam 5.30.0 podem ser configurados com um período de tempo limite de escalonamento automático que permite que as tarefas do Presto

terminem de ser executadas antes que seu nó seja desativado. Para obter mais informações, consulte [Usar a escalabilidade automática do Presto com desativação tranquila](#).

- Criação de instância de frota com nova opção de estratégia de alocação — Uma nova opção de estratégia de alocação está disponível na EMR versão 5.12.1 e posterior. Ele oferece provisionamento de cluster mais rápido, alocação de spot mais precisa e menos interrupção de instâncias spot. Atualizações para funções de EMR serviço não padrão são necessárias. Consulte [Configurar frotas de instâncias](#).
- comandos `sudo systemctl stop` e `sudo systemctl start` — Na EMR versão 5.30.0 e posterior, qual sistema operacional use Amazon Linux 2 usa e comanda para reiniciar serviços. `EMR sudo systemctl stop sudo systemctl start` Para obter mais informações, consulte [Como reinício um serviço na AmazonEMR?](#).

Alterações, melhorias e problemas resolvidos

- EMR versão 5.30.0 não instala o Ganglia por padrão. É possível selecionar explicitamente o Ganglia para ser instalado ao criar um cluster.
- Otimizações do desempenho do Spark
- Otimizações do desempenho do Presto
- O Python 3 é o padrão para a Amazon EMR versão 5.30.0 e posterior.
- O grupo de segurança gerenciado padrão para acesso ao serviço em sub-redes privadas foi atualizado com novas regras. Se você usar um grupo de segurança personalizado para acesso ao serviço, será necessário incluir as mesmas regras do grupo de segurança gerenciado padrão. Para obter mais informações, consulte [Grupo EMR de segurança gerenciado pela Amazon para acesso a serviços \(sub-redes privadas\)](#). Se você usa uma função de serviço personalizada para a AmazonEMR, deve conceder permissão para que ela `ec2:describeSecurityGroups` EMR possa validar se os grupos de segurança foram criados corretamente. Se você usar o `EMR_DefaultRole`, essa permissão já estará incluída na política gerenciada padrão.

Problemas conhecidos

- Limite inferior de “Máximo de arquivos abertos” em versões mais antigas AL2 [corrigido em versões mais recentes]. EMRLançamentos da Amazon: `emr-5.30.x`, `emr-5.31.0`, `emr-5.32.0`, `emr-6.0.0`, `emr-6.1.0` e `emr-6.2.0` são baseados em versões mais antigas do Linux of Amazon 2 (AL2, que têm uma configuração de limite inferior para “Máximo de arquivos abertos” quando os clusters da Amazon são criados com o padrão. EMR AMI As EMR versões 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 e posteriores incluem uma correção permanente com uma configuração maior de “Máximo de arquivos abertos”. Versões com o limite inferior de arquivos abertos causam o erro “Muitos arquivos abertos” ao ser enviado um trabalho do Spark. Nas versões afetadas, o padrão da Amazon AMI tem uma configuração EMR padrão de `ulimit` de 4096 para “Máximo de arquivos abertos”, que é inferior ao limite de 65536 arquivos no Linux 2. latest Amazon AMI A configuração inferior de `ulimit` para “Máximo de arquivos abertos” causa falhas em trabalhos do Spark quando o driver e o executor do Spark tentam abrir mais de 4096 arquivos. Para corrigir o problema, a Amazon EMR tem um script de ação de bootstrap (BA) que ajusta a configuração `ulimit` na criação do cluster.

Se você estiver usando uma EMR versão mais antiga da Amazon que não tem a solução permanente para esse problema, a solução alternativa a seguir permite que você defina explicitamente o `ulimit` do controlador de instância para um máximo de 65536 arquivos.

Defina explicitamente um `ulimit` na linha de comando

1. Edite `/etc/systemd/system/instance-controller.service` para adicionar os seguintes parâmetros à seção Serviço.

```
LimitNOFILE=65536
```

```
LimitNPROC=65536
```

2. Reiniciar InstanceController

```
$ sudo systemctl daemon-reload
```

```
$ sudo systemctl restart instance-controller
```

Defina um `ulimit` usando a ação de bootstrap (BA)

Você também pode usar um script de ação de bootstrap (BA) para configurar o `ulimit instance-controller` para 65536 arquivos na criação do 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
```

- Ajuste de escala gerenciado

As operações de ajuste de escala gerenciado nos clusters das versões 5.30.0 e 5.30.1 sem o Presto instalado podem causar falhas na aplicação ou fazer com que um grupo de instâncias ou uma frota de instâncias uniforme permaneça no estado ARRESTED, sobretudo quando uma operação de redução da escala verticalmente logo é seguida por uma operação de aumento da escala verticalmente.

Como solução alternativa, escolha o Presto como um aplicativo a ser instalado ao criar um cluster com as EMR versões 5.30.0 e 5.30.1 da Amazon, mesmo que seu trabalho não exija o Presto.

- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSH como usuário do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do hadoop.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo `keytab` está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

- O mecanismo de banco de dados padrão para o Hue 4.6.0 é SQLite, o que causa problemas quando você tenta usar o Hue com um banco de dados externo. Para corrigir isso, defina `engine` na sua classificação de configuração `hue.ini` como `mysql`. Esse problema foi corrigido na EMR versão 5.30.1 da Amazon.
- Quando você usa o Spark com a formatação de localização de partições do Hive para ler dados no Amazon S3 e executa o Spark nas EMR versões 5.30.0 a 5.36.0 e 6.2.0 a 6.9.0 da Amazon, você pode encontrar um problema que impede seu cluster de ler os dados corretamente. Isso poderá acontecer se suas partições tiverem todas as características a seguir:
 - Duas ou mais partições são verificadas na mesma tabela.
 - Pelo menos um caminho de diretório de partição é um prefixo de pelo menos outro caminho de diretório de partição, por exemplo, `s3://bucket/table/p=a` é um prefixo de `s3://bucket/table/p=a b`.
 - O primeiro caractere que segue o prefixo no outro diretório de partição tem um valor UTF-8 menor que o `/` caractere (U+002F). Por exemplo, o caractere de espaço (U+0020) que ocorre entre `a` e `b` em `s3://bucket/table/p=a b` se enquadra nessa categoria. Observe que existem 14 outros caracteres que não são de controle: `!"#$%&'()*+,-.` Para obter mais informações, consulte a [tabela de codificação UTF-8 e caracteres Unicode](#).

Como solução alternativa para esse problema, defina a configuração `spark.sql.sources.fastS3PartitionDiscovery.enabled` como `false` na classificação `spark-defaults`.

Versão 5.29.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.29.0 da Amazon. As alterações são referentes à versão 5.28.1.

Data da versão inicial: 17 de janeiro de 2020

Atualizações

- Atualizado para AWS SDK for Java a versão 1.11.682
- Atualizado o Hive para a versão 2.3.6
- Atualizado o Flink para a versão 1.9.1
- Atualizado o EMRFS para a versão 2.38.0
- Conector EMR DynamoDB atualizado para a versão 4.13.0

Alterações, melhorias e problemas resolvidos

- Spark
 - Otimizações do desempenho do Spark
- EMRFS
 - O Guia de gerenciamento é atualizado para as configurações padrão `emrfs-site.xml` para uma visualização consistente.

Problemas conhecidos

- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSH como `hadoop` usuário do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do `hadoop`.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo `keytab` está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

Versão 5.28.1

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.28.1 da Amazon. As alterações são referentes à versão 5.28.0.

Data da versão inicial: 10 de janeiro de 2020

Alterações, melhorias e problemas resolvidos

- Spark
 - Correção de problemas de compatibilidade do Spark.
- CloudWatch Métricas
 - Foi corrigida a publicação do Amazon CloudWatch Metrics em um EMR cluster com vários nós primários.
- Desabilitada mensagem de log
 - Desabilitada mensagem de log falsa, "... uso de versão antiga (<4.5.8) do cliente Apache http".

Problemas conhecidos

- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto

as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSH como usuário do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do hadoop.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo keytab está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

Versão 5.28.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.28.0 da Amazon. As alterações são referentes à 5.27.0.

Data da versão inicial: 12 de novembro de 2019

Atualizações

- Flink atualizado para a versão 1.9.0
- Atualizado o Hive para a versão 2.3.6
- Atualizado MXNet para a versão 1.5.1
- Phoenix atualizado para a versão 4.14.3
- Presto atualizado para a versão 0.227
- Zeppelin atualizado para a versão 0.8.2

Novos atributos

- [O Apache Hudi](#) agora está disponível EMR para a Amazon instalar quando você cria um cluster. Para obter mais informações, consulte [Hudi](#).
- (25 de novembro de 2019) Agora você pode optar por executar várias etapas em paralelo para melhorar a utilização do cluster e economizar custos. Pode também cancelar etapas pendentes e em execução. Para obter mais informações, consulte [Trabalhar com etapas usando o console AWS CLI e](#).
- (3 de dezembro de 2019) Agora você pode criar e executar EMR clusters no AWS Outposts. AWS Outposts habilita AWS serviços, infraestrutura e modelos operacionais nativos em instalações locais. Em AWS Outposts ambientes, você pode usar as mesmas AWS APIs ferramentas e infraestrutura que usa na AWS nuvem. Para obter mais informações, consulte [EMRclusters on AWS Outposts](#).
- (11 de março de 2020) A partir da EMR versão 5.28.0 da Amazon, você pode criar e executar EMR clusters da Amazon em uma sub-rede de Zonas AWS Locais como uma extensão lógica de uma AWS região que suporta Zonas Locais. Uma zona local permite que os EMR recursos da Amazon e um subconjunto de AWS serviços, como serviços de computação e armazenamento, estejam localizados mais perto dos usuários, fornecendo acesso de latência muito baixa a aplicativos executados localmente. Para obter uma lista das zonas locais disponíveis, consulte [Zonas locais da AWS](#). Para obter informações sobre como acessar as Zonas AWS Locais disponíveis, consulte [Regiões, Zonas de Disponibilidade e Zonas Locais](#).

No momento, as Zonas Locais não oferecem suporte a Amazon EMR Notebooks e não oferecem suporte a conexões diretamente com a Amazon EMR usando a interface VPC endpoint (AWS PrivateLink).

Alterações, melhorias e problemas resolvidos

- Suporte expandido do aplicativo para clusters de alta disponibilidade
 - Para obter mais informações, consulte [Aplicativos compatíveis em um EMR cluster com vários nós primários](#) no Amazon EMR Management Guide.
- Spark
 - Otimizações da performance
- Hive
 - Otimizações da performance

- Presto
 - Otimizações da performance

Problemas conhecidos

- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSH como usuário do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do hadoop.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo keytab está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

Versão 5.27.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.27.0 da Amazon. As alterações são referentes à versão 5.26.0.

Data da versão inicial: 23 de setembro de 2019

Atualizações

- AWS SDK for Java 1.11.615
- Flink 1.8.1
- JupyterHub 1.0.0
- Spark 2.4.4
- Tensorflow 1.14.0
- Conectores e drivers:
 - Conector do DynamoDB 4.12.0

Novos atributos

- (24 de outubro de 2019) Os seguintes novos recursos em EMR notebooks estão disponíveis em todos os lançamentos da AmazonEMR.
 - Agora você pode associar repositórios Git a EMR notebooks para armazenar seus cadernos em um ambiente com controle de versão. Você pode compartilhar códigos com pares e reutilizar cadernos Jupyter existentes por meio de repositórios do Git remotos. Para obter mais informações, consulte [Associar repositórios Git aos Amazon EMR Notebooks no Guia de gerenciamento](#) da Amazon. EMR
 - O [utilitário nbdime](#) agora está disponível em EMR notebooks para simplificar a comparação e mesclagem de notebooks.
 - EMRnotebooks agora são compatíveis JupyterLab. JupyterLab é um ambiente de desenvolvimento interativo baseado na Web totalmente compatível com os notebooks Jupyter. Agora você pode optar por abrir seu caderno em qualquer um dos editores de cadernos JupyterLab ou no editor de cadernos Jupyter.
- (30 de outubro de 2019) Com EMR as versões 5.25.0 e posteriores da Amazon, você pode se conectar à interface do usuário do servidor de histórico do Spark na página de resumo do cluster ou na guia Histórico do aplicativo no console. Em vez de configurar um proxy web por meio de uma SSH conexão, você pode acessar rapidamente a interface do servidor de histórico do Spark para visualizar as métricas do aplicativo e acessar arquivos de log relevantes para clusters ativos e encerrados. Para obter mais informações, consulte [Acesso fora do cluster a interfaces de usuário de aplicativos persistentes](#) no Amazon EMR Management Guide.

Alterações, melhorias e problemas resolvidos

- EMRCluster Amazon com vários nós primários
 - Você pode instalar e executar o Flink em um EMR cluster da Amazon com vários nós primários. Para obter mais informações, consulte [Supported applications and features](#).
 - Você pode configurar a criptografia HDFS transparente em um EMR cluster da Amazon com vários nós primários. Para obter mais informações, consulte [Criptografia HDFS transparente em EMR clusters com vários nós primários](#).
 - Agora você pode modificar a configuração dos aplicativos em execução em um EMR cluster da Amazon com vários nós primários. Para obter mais informações, consulte [Supplying a Configuration for an Instance Group in a Running Cluster](#).
- Conector Amazon EMR - DynamoDB
 - O Amazon EMR -DynamoDB Connector agora oferece suporte aos seguintes tipos de dados do DynamoDB: boolean, list, map, item, null. Para obter mais informações, consulte [Configurar uma tabela do Hive para executar comandos do Hive](#).

Problemas conhecidos

- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSH como hadoop usuário do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do hadoop.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo keytab está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

Versão 5.26.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.26.0 da Amazon. As alterações são referentes à versão 5.25.0.

Data da versão inicial: 8 de agosto de 2019

Data da última atualização: 19 de agosto de 2019

Atualizações

- AWS SDK for Java 1.11.595
- HBase1.4.10
- Phoenix 4.14.2
- Conectores e drivers:
 - Conector do DynamoDB 4.11.0
 - Conector do MariaDB 2.4.2
 - Driver Amazon Redshift 1.2.32.1056 JDBC

Novos atributos

- (Beta) Com o Amazon EMR 5.26.0, você pode lançar um cluster que se integra ao Lake Formation. Essa integração fornece acesso refinado em nível de coluna a bancos de dados e tabelas no Glue Data Catalog. AWS Ele também permite o login único federado em EMR notebooks ou Apache Zeppelin a partir de um sistema de identidade corporativo. Para obter mais informações, consulte [Integração da Amazon EMR com AWS Lake Formation \(Beta\)](#).
- (19 de agosto de 2019) O acesso público de EMR bloqueio da Amazon agora está disponível com todos os EMR lançamentos da Amazon que oferecem suporte a grupos de segurança. Bloquear

o acesso público é uma configuração de toda a conta aplicada a cada AWS região. Bloquear o acesso público impede que um cluster seja iniciado quando qualquer grupo de segurança associado ao cluster tem uma regra que permite tráfego de entrada de IPv4 0.0.0.0/0 ou IPv6:: /0 (acesso público) em uma porta, a menos que uma porta seja especificada como uma exceção. A porta 22 é uma exceção por padrão. Para obter mais informações, consulte [Usando o Amazon EMR Block Public Access](#) no Amazon EMR Management Guide.

Alterações, melhorias e problemas resolvidos

- EMRCadernos
 - Com a EMR versão 5.26.0 e versões posteriores, o EMR Notebooks oferece suporte a bibliotecas Python com escopo de notebook, além das bibliotecas Python padrão. Você pode instalar bibliotecas com escopo de caderno de dentro do editor de caderno sem precisar recriar um cluster ou reanexar um caderno a um cluster. As bibliotecas com escopo de caderno são criadas em um ambiente Python virtual para serem aplicadas somente à sessão de caderno atual. Isso permite isolar as dependências do caderno. Para obter mais informações, consulte [Usando bibliotecas com escopo de notebook](#) no Amazon EMR Management Guide.
- EMRFS
 - Você pode ativar um recurso ETag de verificação (Beta) configurando `fs.s3.consistent.metadata.etag.verification.enabled` como `true`. Com esse recurso, EMRFS usa o Amazon S3 ETags para verificar se os objetos que estão sendo lidos são a versão mais recente disponível. Esse recurso é útil para casos de read-after-update uso em que os arquivos no Amazon S3 são sobrescritos, mantendo o mesmo nome. Atualmente, esse recurso de ETag verificação não funciona com o S3 Select. Para obter mais informações, consulte [Configurar visualização consistente](#).
- Spark
 - As seguintes otimizações agora estão habilitadas por padrão: remoção dinâmica de partições, DISTINCT antes, melhorias na inferência de estatísticas do SQL plano para consultas JOIN seguidas INTERSECT, nivelamento de DISTINCT subconsultas escalares, reordenamento otimizado de junções e junção com filtro bloom. Para obter mais informações, consulte [Otimizar a performance do Spark](#).
 - Aprimorada a geração de código de estágio completo para Sort-Merge Join.
 - Aprimorado o fragmento de consulta e a reutilização de subconsultas.
 - Melhorias na pré-alocação de executores na inicialização do Spark.

- As junções com filtro de Bloom não são mais aplicadas quando o lado menor da junção inclui uma dica de transmissão.
- Tez
 - Resolvido um problema com o Tez. O Tez UI agora funciona em um EMR cluster da Amazon com vários nós primários.

Problemas conhecidos

- Os recursos aprimorados de geração de código em todo o estágio de Sort Merge Join podem aumentar a pressão de memória quando habilitados. Essa otimização melhora a performance, mas pode resultar em novas tentativas ou falhas de trabalho se `spark.yarn.executor.memoryOverheadFactor` não for ajustado para fornecer memória suficiente. Para desabilitar esse atributo, defina `spark.sql.sortMergeJoinExec.extendedCodegen.enabled` como falso.
- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSH como `hadoop` usuário do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do `hadoop`.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo `keytab` está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

Versão 5.25.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.25.0 da Amazon. As alterações são referentes à versão 5.24.1.

Data da versão inicial: 17 de julho de 2019

Data da última atualização: 30 de outubro de 2019

Amazon EMR 5.25.0

Atualizações

- AWS SDK for Java 1.11.566
- Hive 2.3.5
- Presto 0.220
- Spark 2.4.3
- TensorFlow 1.13.1
- Tez 0.9.2
- Zookeeper 3.4.14

Novos atributos

- (30 de outubro de 2019) A partir da EMR versão 5.25.0 da Amazon, você pode se conectar à interface do usuário do servidor de histórico do Spark na página de resumo do cluster ou na guia Histórico do aplicativo no console. Em vez de configurar um proxy web por meio de uma SSH conexão, você pode acessar rapidamente a interface do servidor de histórico do Spark para visualizar as métricas do aplicativo e acessar arquivos de log relevantes para clusters ativos e

encerrados. Para obter mais informações, consulte [Acesso fora do cluster a interfaces de usuário de aplicativos persistentes](#) no Amazon EMR Management Guide.

Alterações, melhorias e problemas resolvidos

- Spark
 - Aprimorada a performance de algumas junções usando filtros de Bloom para pré-filtrar as entradas. A otimização é desabilitada por padrão e pode ser habilitada com a definição do parâmetro `spark.sql.bloomFilterJoin.enabled` de configuração do Spark como `true`.
 - Aprimorada a performance do agrupamento por colunas do tipo string.
 - Melhorou a memória padrão do executor Spark e a configuração dos núcleos dos tipos de instância R4 para clusters sem instalação. HBase
 - Resolvido um problema anterior com o atributo de remoção dinâmica de partições, em que a tabela removida precisava estar no lado esquerdo da junção.
 - Aprimorado DISTINCT antes da INTERSECT otimização para ser aplicado a casos adicionais envolvendo aliases.
 - Inferência aprimorada das estatísticas do SQL plano para DISTINCT consultas JOIN seguidas. Essa melhoria é desabilitada por padrão e pode ser habilitada pela definição do parâmetro `spark.sql.statsImprovements.enabled` de configuração do Spark como `true`. Essa otimização é exigida pelo atributo Distinct antes do Intersect e será habilitada automaticamente quando `spark.sql.optimizer.distinctBeforeIntersect.enabled` estiver definido como `true`.
 - Otimizada a ordem de junção com base no tamanho da tabela e nos filtros. Essa otimização é desativada por padrão e pode ser ativada com a definição do parâmetro `spark.sql.optimizer.sizeBasedJoinReorder.enabled` de configuração do Spark como `true`.

Para obter mais informações, consulte [Otimizar a performance do Spark](#).

- EMRFS
 - A EMRFS configuração, `fs.s3.buckets.create.enabled`, agora está desativada por padrão. Por meio de testes, descobrimos que a desabilitação dessa configuração melhora a performance e evita a criação não intencional de buckets do S3. Se sua aplicação depende dessa funcionalidade, você pode habilitá-la definindo a propriedade `fs.s3.buckets.create.enabled` como `true` na classificação de configuração `emrfs-site`. Para obter informações, consulte [Supplying a Configuration when Creating a Cluster](#).

- Melhorias na criptografia de disco local e na criptografia do S3 nas configurações de segurança (5 de agosto de 2019)
 - Separadas as configurações de criptografia do Amazon S3 das configurações de criptografia de disco local na configuração de segurança.
 - Foi adicionada uma opção para ativar a EBS criptografia com a versão 5.24.0 e posterior. Selecionar essa opção criptografa o volume do dispositivo raiz, além dos volumes de armazenamento. As versões anteriores exigiam o uso de um recurso personalizado AMI para criptografar o volume do dispositivo raiz.
 - Para obter mais informações, consulte [Opções de criptografia](#) no Amazon EMR Management Guide.

Problemas conhecidos

- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSH como hadoop usuário do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do hadoop.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo keytab está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas

definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

Versão 5.24.1

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.24.1 da Amazon. As alterações são referentes à versão 5.24.0.

Data da versão inicial: 26 de junho de 2019

Alterações, melhorias e problemas resolvidos

- Atualizou o Amazon Linux padrão AMI para Amazon EMR para incluir atualizações importantes de segurança do kernel Linux, incluindo o problema de TCP SACK negação de serviço ([AWS-2019-005](#)).

Problemas conhecidos

- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSH como hadoop usuário do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do hadoop.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo keytab está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

Versão 5.24.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.24.0 da Amazon. As alterações são referentes à versão 5.23.0.

Data da versão inicial: 11 de junho de 2019

Data da última atualização: 5 de agosto de 2019

Atualizações

- 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 1.11.546
- Conectores e drivers:
 - Conector do DynamoDB 4.9.0
 - Conector do MariaDB 2.4.1
 - Driver Amazon Redshift 1.2.27.1051 JDBC

Alterações, melhorias e problemas resolvidos

- Spark

- Adicionada otimização para remover partições dinamicamente. A otimização está desabilitada por padrão. Para habilitá-la, defina o parâmetro `spark.sql.dynamicPartitionPruning.enabled` de configuração do Spark como `true`.
- Aprimorada a performance de consultas INTERSECT. Essa otimização está desabilitada por padrão. Para habilitá-la, defina o parâmetro `spark.sql.optimizer.distinctBeforeIntersect.enabled` de configuração do Spark como `true`.
- Adicionada otimização para nivelar subconsultas escalares com agregados que usam a mesma relação. A otimização está desabilitada por padrão. Para habilitá-la, defina o parâmetro `spark.sql.optimizer.flattenScalarSubqueriesWithAggregates.enabled` de configuração do Spark como `true`.
- Aprimorada a geração de código em todo o estágio.

Para obter mais informações, consulte [Otimizar a performance do Spark](#).

- Melhorias na criptografia de disco local e na criptografia do S3 nas configurações de segurança (5 de agosto de 2019)
 - Separadas as configurações de criptografia do Amazon S3 das configurações de criptografia de disco local na configuração de segurança.
 - Foi adicionada uma opção para ativar a EBS criptografia. Selecionar essa opção criptografa o volume do dispositivo raiz, além dos volumes de armazenamento. As versões anteriores exigiam o uso de um recurso personalizado AMI para criptografar o volume do dispositivo raiz.
 - Para obter mais informações, consulte [Opções de criptografia](#) no Amazon EMR Management Guide.

Problemas conhecidos

- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSH como usuário do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do hadoop.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo keytab está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

Versão 5.23.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.23.0 da Amazon. As alterações são referentes à versão 5.22.0.

Data da versão inicial: 1.º de abril de 2019

Data da última atualização: 30 de abril de 2019

Atualizações

- AWS SDK for Java 1.11.519

Novos atributos

- (30 de abril de 2019) Com o Amazon EMR 5.23.0 e versões posteriores, você pode iniciar um cluster com três nós principais para oferecer suporte à alta disponibilidade de aplicativos como YARN Resource Manager, Spark HDFS NameNode, Hive e Ganglia. O nó primário não é mais um possível ponto de falha único com esse recurso. Se um dos nós primários falhar, a Amazon EMR automaticamente passa para um nó primário em espera e substitui o nó primário com falha por um

novo com as mesmas ações de configuração e bootstrap. Para obter mais informações, consulte [Plan and Configure Primary Nodes](#).

Problemas conhecidos

- Tez UI (corrigido na EMR versão 5.26.0 da Amazon)

O Tez UI não funciona em um EMR cluster com vários nós primários.

- Hue (corrigido na EMR versão 5.24.0 da Amazon)
 - O Hue executado na Amazon EMR não oferece suporte ao Solr. A partir da EMR versão 5.20.0 da Amazon, um problema de configuração incorreta faz com que o Solr seja ativado e uma mensagem de erro inofensiva apareça semelhante à seguinte:

```
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',))
```

Para evitar que a mensagem de erro do Solr seja exibida:

1. Conecte-se à linha de comando do nó primário usando SSH o.
2. Use um editor de texto para abrir o arquivo `hue.ini`. Por exemplo:

```
sudo vim /etc/hue/conf/hue.ini
```

3. Pesquise o termo `appblacklist` e modifique a linha para o seguinte:

```
appblacklist = search
```

4. Salve as alterações e reinicie o Hue, conforme mostrado no exemplo a seguir:

```
sudo stop hue; sudo start hue
```

- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo.

O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSH como hadoop usuário do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do hadoop.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo keytab está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

Versão 5.22.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.22.0 da Amazon. As alterações são referentes à versão 5.21.0.

Important

A partir da EMR versão 5.22.0 da Amazon, a Amazon EMR usa o AWS Signature versão 4 exclusivamente para autenticar solicitações para o Amazon S3. As EMR versões anteriores da Amazon usam a AWS Signature Version 2 em alguns casos, a menos que as notas de lançamento indiquem que a Signature Version 4 é usada exclusivamente. Para obter mais informações, consulte [Autenticação de solicitações \(AWS assinatura versão 4\)](#) e [Solicitações de autenticação \(AWS assinatura versão 2\) no Guia](#) do desenvolvedor do Amazon Simple Storage Service.

Data da versão inicial: 20 de março de 2019

Atualizações

- Flink 1.7.1
- HBase1.4.9
- Oozie 5.1.0
- Phoenix 4.14.1
- Zeppelin 0.8.1
- Conectores e drivers:
 - Conector do DynamoDB 4.8.0
 - Conector do MariaDB 2.2.6
 - Driver Amazon Redshift 1.2.20.1043 JDBC

Novos atributos

- Modificou a EBS configuração padrão para tipos de EC2 instância com armazenamento EBS somente. Quando você cria um cluster usando a EMR versão 5.22.0 e posterior da Amazon, a quantidade padrão de EBS armazenamento aumenta com base no tamanho da instância. Além disso, dividimos o aumento do armazenamento em vários volumes, proporcionando maior IOPS desempenho. Se quiser usar uma configuração de armazenamento de EBS instância diferente, você pode especificá-la ao criar um EMR cluster ou adicionar nós a um cluster existente. Para obter mais informações sobre a quantidade de armazenamento e o número de volumes alocados por padrão para cada tipo de instância, consulte [EBSArmazenamento padrão para instâncias](#) no Amazon EMR Management Guide.

Alterações, melhorias e problemas resolvidos

- Spark
 - Introduziu uma nova propriedade de configuração para o Spark emYARN, `spark.yarn.executor.memoryOverheadFactor`. O valor dessa propriedade é um fator de escala que define o valor da sobrecarga de memória como uma porcentagem da memória do executor, com um mínimo de 384 MB. Se a sobrecarga de memória for definida explicitamente usando `spark.yarn.executor.memoryOverhead`, essa propriedade não terá efeito. O valor padrão é `0.1875`, representando 18,75%. Esse padrão para a Amazon

EMR deixa mais espaço nos YARN contêineres para sobrecarga de memória do executor do que o padrão de 10% definido internamente pelo Spark. O EMR padrão da Amazon de 18,75% mostrou empiricamente menos falhas relacionadas à memória nos benchmarks -DS. TPC

- Backport [SPARK-26316](#) para melhorar o desempenho.
- Nas EMR versões 5.19.0, 5.20.0 e 5.21.0 da Amazon, os rótulos dos YARN nós são armazenados em um diretório. HDFS Em algumas situações, isso leva a atrasos na inicialização do nó central causando, em seguida, tempo limite do cluster e falha na inicialização. A partir do Amazon EMR 5.22.0, esse problema foi resolvido. YARNos rótulos dos nós são armazenados no disco local de cada nó do cluster, evitando dependências emHDFS.

Problemas conhecidos

- Hue (corrigido na EMR versão 5.24.0 da Amazon)
 - O Hue executado na Amazon EMR não oferece suporte ao Solr. A partir da EMR versão 5.20.0 da Amazon, um problema de configuração incorreta faz com que o Solr seja ativado e uma mensagem de erro inofensiva apareça semelhante à seguinte:

```
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',))
```

Para evitar que a mensagem de erro do Solr seja exibida:

1. Conecte-se à linha de comando do nó primário usando SSH o.
2. Use um editor de texto para abrir o arquivo `hue.ini`. Por exemplo:

```
sudo vim /etc/hue/conf/hue.ini
```

3. Pesquise o termo `appblacklist` e modifique a linha para o seguinte:

```
appblacklist = search
```

4. Salve as alterações e reinicie o Hue, conforme mostrado no exemplo a seguir:

```
sudo stop hue; sudo start hue
```


- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSH como usuário do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do hadoop.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo keytab está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

Versão 5.21.1

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.21.1 da Amazon. As alterações são referentes à versão 5.21.0.

Data da versão inicial: 18 de julho de 2019

Alterações, melhorias e problemas resolvidos

- Atualizou o Amazon Linux padrão AMI para Amazon EMR para incluir atualizações importantes de segurança do kernel Linux, incluindo o problema de TCP SACK negação de serviço ([AWS-2019-005](#)).

Problemas conhecidos

- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSH como `hadoop` usuário do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do `hadoop`.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo `keytab` está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

Versão 5.21.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.21.0 da Amazon. As alterações são referentes à versão 5.20.0.

Data da versão inicial: 18 de fevereiro de 2019

Data da última atualização: 3 de abril de 2019

Atualizações

- Flink 1.7.0
- Presto 0.215
- AWS SDK for Java 1.11.479

Novos atributos

- (3 de abril de 2019) Com a Amazon EMR versão 5.21.0 e posterior, você pode substituir as configurações do cluster e especificar classificações de configuração adicionais para cada grupo de instâncias em um cluster em execução. Você faz isso usando o EMR console da Amazon, o AWS Command Line Interface (AWS CLI) ou AWS SDK o. Para obter mais informações, consulte [Supplying a Configuration for an Instance Group in a Running Cluster](#).

Alterações, melhorias e problemas resolvidos

- Zeppelin
 - Porta traseira [ZEPPELIN-3878](#).

Problemas conhecidos

- Hue (corrigido na EMR versão 5.24.0 da Amazon)
 - O Hue executado na Amazon EMR não oferece suporte ao Solr. A partir da EMR versão 5.20.0 da Amazon, um problema de configuração incorreta faz com que o Solr seja ativado e uma mensagem de erro inofensiva apareça semelhante à seguinte:

```
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',))
```

Para evitar que a mensagem de erro do Solr seja exibida:

1. Conecte-se à linha de comando do nó primário usando SSH o.
2. Use um editor de texto para abrir o arquivo `hue.ini`. Por exemplo:

```
sudo vim /etc/hue/conf/hue.ini
```

3. Pesquise o termo `appblacklist` e modifique a linha para o seguinte:

```
appblacklist = search
```

4. Salve as alterações e reinicie o Hue, conforme mostrado no exemplo a seguir:

```
sudo stop hue; sudo start hue
```

- Tez
- Esse problema foi corrigido no Amazon EMR 5.22.0.

Quando você se conecta à interface do usuário do Tez em `http://MasterDNS:8080/tez-ui` por meio de uma SSH conexão com o nó primário do cluster, surge o erro “Falha na operação do adaptador - Timeline server () está fora de alcance. ATS “Está inativo ou não CORS está ativado” aparece ou as tarefas mostram inesperadamente N/A.

Isso é causado pela interface do usuário do Tez fazendo solicitações ao YARN Timeline Server usando `localhost` em vez do nome do host, o nó primário. Como solução alternativa, um script está disponível para execução como ação ou etapa de bootstrap. O script atualiza o nome do host no arquivo `configs.env` do Tez. Para obter mais informações e a localização do script, consulte [Instruções de bootstrap](#).

- Nas EMR versões 5.19.0, 5.20.0 e 5.21.0 da Amazon, os rótulos dos YARN nós são armazenados em um diretório. HDFS Em algumas situações, isso leva a atrasos na inicialização do nó central causando, em seguida, tempo limite do cluster e falha na inicialização. A partir do Amazon EMR 5.22.0, esse problema foi resolvido. YARNos rótulos dos nós são armazenados no disco local de cada nó do cluster, evitando dependências emHDFS.
- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSH como hadoop usuário do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do hadoop.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo keytab está localizado em `/etc/hadoop.keytab` e a entidade principal está na forma de `hadoop/<hostname>@<REALM>`.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

Versão 5.20.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.20.0 da Amazon. As alterações são referentes à versão 5.19.0.

Data da versão inicial: 18 de dezembro de 2018

Data da última atualização: 22 de janeiro de 2019

Atualizações

- 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

Novos atributos

- (22 de janeiro de 2019) O Kerberos na Amazon EMR foi aprimorado para oferecer suporte à autenticação de diretores externos. KDC Isso centraliza o gerenciamento principal porque vários clusters podem compartilhar um único externoKDC. Além disso, o externo KDC pode ter uma relação de confiança entre regiões com um domínio do Active Directory. Isso permite que todos os clusters autentiquem principais do Active Directory. Para obter mais informações, consulte [Usar a autenticação Kerberos](#) no Amazon EMR Management Guide.

Alterações, melhorias e problemas resolvidos

- Amazon Linux padrão AMI para Amazon EMR
 - O pacote Python3 foi atualizado do python 3.4 para 3.6.
- O committer EMRFS otimizado para S3
 - O committer EMRFS otimizado para S3 agora está habilitado por padrão, o que melhora o desempenho de gravação. Para obter mais informações, consulte [Use o committer EMRFS otimizado para S3](#).
- Hive
 - Portado para trás [HIVE-16686](#).
- Glue com Spark e Hive
 - Na EMR versão 5.20.0 ou posterior, a remoção paralela de partições é ativada automaticamente para Spark e Hive quando o AWS Glue Data Catalog é usado como metastore. Essa alteração reduz significativamente o tempo de planejamento de consultas ao executar várias solicitações em paralelo para recuperar partições. O número total de segmentos que podem ser executados simultaneamente varia entre 1 e 10. O valor padrão é 5, que é uma configuração recomendada.

Você pode alterá-lo especificando a propriedade `aws.glue.partition.num.segments` na classificação de configuração `hive-site`. Se ocorrer controle de utilização, você poderá desativar o atributo alterando o valor para 1. Para obter mais informações, consulte a [Estrutura de segmentos do AWS Glue](#).

Problemas conhecidos

- Hue (corrigido na EMR versão 5.24.0 da Amazon)
 - O Hue executado na Amazon EMR não oferece suporte ao Solr. A partir da EMR versão 5.20.0 da Amazon, um problema de configuração incorreta faz com que o Solr seja ativado e uma mensagem de erro inofensiva apareça semelhante à seguinte:

```
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',))
```

Para evitar que a mensagem de erro do Solr seja exibida:

1. Conecte-se à linha de comando do nó primário usando SSH o.
2. Use um editor de texto para abrir o arquivo `hue.ini`. Por exemplo:

```
sudo vim /etc/hue/conf/hue.ini
```

3. Pesquise o termo `appblacklist` e modifique a linha para o seguinte:

```
appblacklist = search
```

4. Salve as alterações e reinicie o Hue, conforme mostrado no exemplo a seguir:

```
sudo stop hue; sudo start hue
```

- Tez
 - Esse problema foi corrigido no Amazon EMR 5.22.0.

Quando você se conecta à interface do usuário do Tez em `http://MasterDNS:8080/tez-ui` por meio de uma SSH conexão com o nó primário do cluster, surge o erro “Falha na operação

do adaptador - Timeline server () está fora de alcance. ATS “Está inativo ou não CORS está ativado” aparece ou as tarefas mostram inesperadamente N/A.

Isso é causado pela interface do usuário do Tez fazendo solicitações ao YARN Timeline Server usando, localhost em vez do nome do host, o nó primário. Como solução alternativa, um script está disponível para execução como ação ou etapa de bootstrap. O script atualiza o nome do host no arquivo configs.env do Tez. Para obter mais informações e a localização do script, consulte [Instruções de bootstrap](#).

- Nas EMR versões 5.19.0, 5.20.0 e 5.21.0 da Amazon, os rótulos dos YARN nós são armazenados em um diretório. HDFS Em algumas situações, isso leva a atrasos na inicialização do nó central causando, em seguida, tempo limite do cluster e falha na inicialização. A partir do Amazon EMR 5.22.0, esse problema foi resolvido. YARNos rótulos dos nós são armazenados no disco local de cada nó do cluster, evitando dependências emHDFS.
- Problema conhecido em clusters com vários nós primários e autenticação Kerberos

Se você executar clusters com vários nós primários e autenticação Kerberos nas EMR versões 5.20.0 e posteriores da Amazon, poderá encontrar problemas com operações de cluster, como redução de escala ou envio de etapas, depois que o cluster estiver em execução por algum tempo. O período depende do período de validade do tíquete do Kerberos que você definiu. O problema de redução da escala verticalmente afeta tanto as solicitações de redução automática quanto as de reduções explícitas que você enviou. Operações adicionais de cluster também podem ser afetadas.

Solução:

- SSHcomo hadoop usuário do nó primário principal do EMR cluster com vários nós primários.
- Execute o comando a seguir para renovar o tíquete do Kerberos para o usuário do hadoop.

```
kinit -kt <keytab_file> <principal>
```

Normalmente, o arquivo keytab está localizado em /etc/hadoop.keytab e a entidade principal está na forma de hadoop/<hostname>@<REALM>.

Note

Essa solução alternativa entrará em vigor durante o período de validade do tíquete do Kerberos. Essa duração é de 10 horas por padrão, mas pode ser configurada pelas

definições do Kerberos. Você deve executar novamente o comando acima quando o tíquete do Kerberos expirar.

Versão 5.19.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.19.0 da Amazon. As alterações são referentes à versão 5.18.0.

Data da versão inicial: 7 de novembro de 2018

Data da última atualização: 19 de novembro de 2018

Atualizações

- 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 1.11.433

Novos atributos

- (19 de novembro de 2018) O EMR Notebooks é um ambiente gerenciado baseado no Jupyter Notebook. Ele suporta os kernels mágicos do Spark para PySpark, Spark, SQL Spark R e Scala. EMROs notebooks podem ser usados com clusters criados usando a EMR versão 5.18.0 e posterior da Amazon. Para obter mais informações, consulte [Usando EMR notebooks](#) no Guia de EMR gerenciamento da Amazon.
- O committer EMRFS otimizado para S3 está disponível ao gravar arquivos Parquet usando o Spark e. EMRFS Esse confirmador melhora a performance de gravação. Para obter mais informações, consulte [Use o committer EMRFS otimizado para S3](#).

Alterações, melhorias e problemas resolvidos

- YARN
 - Modificada a lógica que limita o processo mestre da aplicação à execução nos nós centrais. Essa funcionalidade agora usa o recurso de rótulos de YARN nós e as propriedades nas classificações `yarn-site` e `capacity-scheduler` de configuração. Para obter mais informações, consulte <https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-plan-instances-guidelines.html#emr-plan-spot-YARN>.
- Amazon Linux padrão AMI para Amazon EMR
 - `ruby18`, `php56`, e `gcc48` não são mais instalados por padrão. Eles podem ser instalados, se desejado, usando `yum`.
 - A gem do `ruby aws-sdk` não é mais instalada por padrão. Ela pode ser instalada usando `gem install aws-sdk`, se desejado. Componentes específicos também podem ser instalados. Por exemplo, `gem install aws-sdk-s3`.

Problemas conhecidos

- EMRNotebooks — Em algumas circunstâncias, com vários editores de notebook abertos, o editor de notebook pode parecer incapaz de se conectar ao cluster. Se isso acontecer, limpe os cookies do navegador e reabra os editores de cadernos.
- CloudWatch ContainerPending Escalabilidade métrica e automática — (corrigida na versão 5.20.0) A Amazon EMR pode emitir um valor negativo para `ContainerPending`. Se `ContainerPending` for usado em uma regra de escalabilidade automática, a escalabilidade automática não se comportará conforme esperado. Evite usar `ContainerPending` com escalabilidade automática.
- Nas EMR versões 5.19.0, 5.20.0 e 5.21.0 da Amazon, os rótulos dos YARN nós são armazenados em um diretório HDFS. Em algumas situações, isso leva a atrasos na inicialização do nó central causando, em seguida, tempo limite do cluster e falha na inicialização. A partir do Amazon EMR 5.22.0, esse problema foi resolvido. YARN os rótulos dos nós são armazenados no disco local de cada nó do cluster, evitando dependências em HDFS.

Versão 5.18.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.18.0 da Amazon. As alterações são referentes à versão 5.17.0.

Data da versão inicial: 24 de outubro de 2018

Atualizações

- Flink 1.6.0
- HBase1.4.7
- Presto 0.210
- Spark 2.3.2
- Zeppelin 0.8.0

Novos atributos

- A partir do Amazon EMR 5.18.0, você pode usar o repositório de EMR artefatos da Amazon para criar seu código de trabalho com base nas versões exatas das bibliotecas e dependências que estão disponíveis em versões específicas da Amazon. Para obter mais informações, consulte [Verificar dependências usando o repositório de artefatos do Amazon EMR](#).

Alterações, melhorias e problemas resolvidos

- Hive
 - Adicionado suporte para o S3 Select. Para obter mais informações, consulte [Usar o S3 Select com o Hive para melhorar a performance](#).
- Presto
 - Adicionado suporte para o [S3 Select](#) Pushdown. Para obter mais informações, consulte [Usar S3 Select Pushdown com o Presto para melhorar a performance](#).
- Spark
 - A configuração log4j padrão do Spark foi alterada para lançar logs de contêineres por hora para trabalhos de streaming do Spark. Isso ajuda a evitar a exclusão de logs de trabalhos de streaming do Spark de execução prolongada.

Versão 5.17.1

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.17.1 da Amazon. As alterações são referentes à versão 5.17.0.

Data da versão inicial: 18 de julho de 2019

Alterações, melhorias e problemas resolvidos

- Atualizou o Amazon Linux padrão AMI para Amazon EMR para incluir atualizações importantes de segurança do kernel Linux, incluindo o problema de TCP SACK negação de serviço ([AWS-2019-005](#)).

Versão 5.17.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.17.0 da Amazon. As alterações são referentes à versão 5.16.0.

Data da versão inicial: 30 de agosto de 2018

Atualizações

- Flink 1.5.2
- HBase1.4.6
- Presto 0.206

Novos atributos

- Adicionado suporte para Tensorflow. Para obter mais informações, consulte [TensorFlow](#).

Alterações, melhorias e problemas resolvidos

- JupyterHub
 - Adicionado suporte para a persistência de cadernos no Amazon S3. Para obter mais informações, consulte [Configurar a persistência de cadernos no Amazon S3](#).
- Spark
 - Adicionado suporte para o [S3 Select](#). Para obter mais informações, consulte [Usar o S3 Select com Spark para melhorar a performance das consultas](#).
- Resolveu os problemas com as métricas do Cloudwatch e o recurso de escalabilidade automática na Amazon EMR versão 5.14.0, 5.15.0 ou 5.16.0.

Problemas conhecidos

- Quando você cria um cluster kerberizado com o Livy instalado, o Livy apresenta falha com um erro em que a autenticação simples não está habilitada. A reinicialização do servidor do Livy resolve o problema. Como solução alternativa, adicione uma etapa durante a criação do cluster que execute `sudo restart livy-server` no nó primário.
- Se você usa um Amazon Linux personalizado AMI baseado em um Amazon Linux AMI com uma data de criação de 11/08/2018, o servidor Oozie falhará ao iniciar. Se você usa o Oozie, crie um personalizado AMI com base em um Amazon Linux AMI ID com uma data de criação diferente. Você pode usar o AWS CLI comando a seguir para retornar uma lista de imagens IDs para todo o HVM Amazon Linux AMIs com uma versão 2018.03, junto com a data de lançamento, para que você possa escolher um Amazon Linux apropriado AMI como base. MyRegion Substitua pelo seu identificador de região, como 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
```

Versão 5.16.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.16.0 da Amazon. As alterações são referentes à versão 5.15.0.

Data da versão inicial: 19 de julho de 2018

Atualizações

- 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 1.11.336

- CUDA9.2
- Controlador Redshift 1.2.15.1025 JDBC

Alterações, melhorias e problemas resolvidos

- HBase
 - [Porta traseira -20723 HBASE](#)
- Presto
 - Alterações na configuração para oferecer suporte à LDAP autenticação. Para obter mais informações, consulte [Usando a LDAP autenticação para o Presto na Amazon EMR](#).
- Spark
 - [A versão 2.3.1 do Apache Spark, disponível a partir da EMR versão 5.16.0 da Amazon, endereça -2018-8024 e -2018-1334. CVE CVE](#) Recomendamos que você migre as versões anteriores do Spark para a versão 2.3.1 ou posteriores.

Problemas conhecidos

- Essa versão não é compatível com os tipos de instância c1.medium ou m1.small. Os clusters que usam qualquer um desses tipos de instância não são iniciados. Como solução alternativa, especifique um tipo de instância diferente ou use uma versão diferente.
- Quando você cria um cluster kerberizado com o Livy instalado, o Livy apresenta falha com um erro em que a autenticação simples não está habilitada. A reinicialização do servidor do Livy resolve o problema. Como solução alternativa, adicione uma etapa durante a criação do cluster que execute `sudo restart livy-server` no nó primário.
- Depois que o nó primário for reinicializado ou o controlador de instância for reiniciado, as CloudWatch métricas não serão coletadas e o recurso de escalabilidade automática não estará disponível nas EMR versões 5.14.0, 5.15.0 ou 5.16.0 da Amazon. Esse problema foi corrigido no Amazon EMR 5.17.0.

Versão 5.15.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.15.0 da Amazon. As alterações são referentes à versão 5.14.0.

Data da versão inicial: 21 de junho de 2018

Atualizações

- Atualizado HBase para 1.4.4
- Atualizado Hive para 2.3.3
- Atualizado Hue para 4.2.0
- Atualizado Oozie para 5.0.0
- Atualizado Zookeeper para 3.4.12
- Atualizado para AWS SDK 1.11.333

Alterações, melhorias e problemas resolvidos

- Hive
 - [Portado para trás - 18069 HIVE](#)
- Hue
 - Atualizado o Hue para se autenticar corretamente com o Livy quando o Kerberos está habilitado. Agora, o Livy é suportado ao usar o Kerberos com a Amazon. EMR
- JupyterHub
 - Atualizado JupyterHub para que a Amazon EMR instale bibliotecas de LDAP clientes por padrão.
 - Corrigido um erro no script que gera certificados autoassinados.

Problemas conhecidos

- Essa versão não é compatível com os tipos de instância c1.medium ou m1.small. Os clusters que usam qualquer um desses tipos de instância não são iniciados. Como solução alternativa, especifique um tipo de instância diferente ou use uma versão diferente.
- Depois que o nó primário for reinicializado ou o controlador de instância for reiniciado, as CloudWatch métricas não serão coletadas e o recurso de escalabilidade automática não estará disponível nas EMR versões 5.14.0, 5.15.0 ou 5.16.0 da Amazon. Esse problema foi corrigido no Amazon EMR 5.17.0.

Versão 5.14.1

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.14.1 da Amazon. As alterações são referentes à versão 5.14.0.

Data da versão inicial: 17 de outubro de 2018

Atualizou o padrão da Amazon AMI EMR para resolver possíveis vulnerabilidades de segurança.

Versão 5.14.0

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.14.0 da Amazon. As alterações são referentes à versão 5.13.0.

Data da versão inicial: 4 de junho de 2018

Atualizações

- Atualizado Apache Flink para 1.4.2
- Apache MXnet atualizado para 1.1.0
- Atualizado Apache Sqoop para 1.4.7

Novos atributos

- JupyterHub Suporte adicionado. Para obter mais informações, consulte [JupyterHub](#).

Alterações, melhorias e problemas resolvidos

- EMRFS
 - A userAgent sequência de caracteres nas solicitações para o Amazon S3 foi atualizada para conter as informações de usuário e grupo do principal invocador. Isso pode ser usado com AWS CloudTrail registros para um rastreamento de solicitações mais abrangente.
- HBase
 - Inclui [HBASE-20447](#), que soluciona um problema que pode causar problemas de cache, especialmente com regiões divididas.
- MXnet
 - Adicionadas bibliotecas OpenCV.
- Spark

- Quando o Spark grava arquivos Parquet em um local do Amazon S3 EMRFS usando, FileOutputCommitter o algoritmo foi atualizado para usar a versão 2 em vez da versão 1. Isso reduz o número de renomeações, o que melhora a performance da aplicação. Essa alteração não afeta:
 - Aplicações diferentes do Spark.
 - Aplicativos que gravam em outros sistemas de arquivos, como HDFS (que ainda usam a versão 1 do FileOutputCommitter).
 - Aplicativos que usam outros formatos de saída, como texto ou csv, que já usam gravação EMRFS direta.

Problemas conhecidos

- JupyterHub
 - O uso de classificações de configuração para configurar JupyterHub notebooks Jupyter individuais ao criar um cluster não é suportado. Edite manualmente o arquivo `jupyterhub_config.py` e os arquivos `jupyter_notebook_config.py` para cada usuário. Para obter mais informações, consulte [Configurando JupyterHub](#).
 - JupyterHub falha ao iniciar em clusters dentro de uma sub-rede privada, falhando com a mensagem. `Error: ENOENT: no such file or directory, open '/etc/jupyter/conf/server.crt'` Isso é causado por um erro no script que gera certificados autoassinados. Use a solução alternativa a seguir para gerar certificados autoassinados. Todos os comandos são executados enquanto estão conectados ao nó primário.
 1. Copie o script de geração de certificados do contêiner para o nó primário:

```
sudo docker cp jupyterhub:/tmp/gen_self_signed_cert.sh ./
```

2. Use um editor de texto para alterar a linha 23 e mudar o nome de host público para o nome de host local, conforme mostrado abaixo:

```
local hostname=$(curl -s $EC2_METADATA_SERVICE_URI/local-hostname)
```

3. Execute o script para gerar certificados autoassinados:

```
sudo bash ./gen_self_signed_cert.sh
```

4. Mova os arquivos de certificado que o script gera para o diretório `/etc/jupyter/conf/`:

```
sudo mv /tmp/server.crt /tmp/server.key /etc/jupyter/conf/
```

Você pode acessar `tail` o `jupyter.log` arquivo para verificar se ele JupyterHub foi reiniciado e está retornando um código de resposta 200. Por exemplo:

```
tail -f /var/log/jupyter/jupyter.log
```

Essa ação deve retornar uma resposta semelhante à seguinte:

```
# [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
```

- Depois que o nó primário for reinicializado ou o controlador de instância for reiniciado, as CloudWatch métricas não serão coletadas e o recurso de escalabilidade automática não estará disponível nas EMR versões 5.14.0, 5.15.0 ou 5.16.0 da Amazon. Esse problema foi corrigido no Amazon EMR 5.17.0.

Versão 5.13.0

As notas de lançamento a seguir incluem informações para a EMR versão 5.13.0 da Amazon. As alterações são referentes à versão 5.12.0.

Atualizações

- Atualizado Spark para 2.3.0
- Atualizado HBase para 1.4.2
- Atualizado Presto para 0.194
- Atualizado para AWS SDK for Java 1.11.297

Alterações, melhorias e problemas resolvidos

- Hive
 - Portado para trás [HIVE-15436](#). Hive aprimorado APIs para retornar somente visualizações.

Problemas conhecidos

- MXNet atualmente não tem bibliotecas OpenCV.

Versão 5.12.2

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.12.2 da Amazon. As alterações são referentes à versão 5.12.1.

Data da versão inicial: 29 de agosto de 2018

Alterações, melhorias e problemas resolvidos

- Esta versão aborda uma possível vulnerabilidade de segurança.

Versão 5.12.1

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.12.1 da Amazon. As alterações são referentes à versão 5.12.0.

Data da versão inicial: 29 de março de 2018

Alterações, melhorias e problemas resolvidos

- Atualizou o kernel Amazon Linux do default Amazon Linux AMI para Amazon EMR para resolver possíveis vulnerabilidades.

Versão 5.12.0

As notas de lançamento a seguir incluem informações para a EMR versão 5.12.0 da Amazon. As alterações são referentes à versão 5.11.1.

Atualizações

- AWS SDK para Java 1.11.238 ⇒ 1.11.267. Para obter mais informações, consulte [AWS SDKo log de alterações do Java](#) GitHub.
- Hadoop 2.7.3 ⇒ 2.8.3. Para obter mais informações, consulte [Versões do Apache Hadoop](#).
- Flink 1.3.2 ⇒ 1.4.0. Para obter mais informações, consulte o [Anúncio de versão do Apache Flink 1.4.0](#).

- HBase1.3.1 ⇒ 1.4.0. Para obter mais informações, consulte o [anúncio HBase de lançamento](#).
- Hue 4.0.1 ⇒ 4.1.0. Para obter mais informações, consulte as [Notas de versão](#).
- MxNet 0,12,0 ⇒ 1,0,0. Para obter mais informações, consulte o [MXNetChange Log](#) on GitHub.
- Presto 0.187 ⇒ 0.188. Para obter mais informações, consulte as [Notas de versão](#).

Alterações, melhorias e problemas resolvidos

- Hadoop
 - A propriedade `yarn.resourcemanager.decommissioning.timeout` foi alterada para `yarn.resourcemanager.nodemanager-graceful-decommission-timeout-secs`. Você pode usar essa propriedade para personalizar a redução da escala do cluster verticalmente. Para obter mais informações, consulte [Cluster Scale-Down](#) no Amazon EMR Management Guide.
 - O Hadoop CLI adicionou a `-d` opção ao comando `cp` (copiar), que especifica a cópia direta. Você pode usar isso para evitar a criação de um arquivo `.COPYING` intermediário, o que torna mais rápida a cópia de dados entre o Amazon S3. Para obter mais informações, consulte [HADOOP-12384](#).
- Pig
 - Adicionada a classificação de configuração `pig-env`, que simplifica a configuração das propriedades do ambiente do Pig. Para obter mais informações, consulte [Configurar aplicações](#).
- Presto
 - Adicionada a classificação de configuração `presto-connector-redshift`, que pode ser usada para configurar os valores no arquivo de configuração `redshift.properties` do Presto. Para obter mais informações, consulte [Conector do Redshift](#) na documentação do Presto e [Configurar aplicações](#).
 - O suporte do Presto para EMRFS foi adicionado e é a configuração padrão. As EMR versões anteriores da Amazon usavam o `PrestOS3FileSystem`, que era a única opção. Para obter mais informações, consulte [EMRFS e configuração do PrestOS3 FileSystem](#).

Note

Se você consultar dados subjacentes no Amazon S3 com a EMR versão 5.12.0 da Amazon, podem ocorrer erros do Presto. Isso acontece porque o Presto não consegue obter valores de classificação de configuração em `emrfs-site.xml`. Como solução alternativa, crie um subdiretório `emrfs` em `usr/lib/presto/plugin/hive-`

```
hadoop2/ e crie um link simbólico em usr/lib/presto/plugin/hive-hadoop2/
emrfs para o arquivo /usr/share/aws/emr/emrfs/conf/emrfs-site.xml
existente. Em seguida, reinicie o processo presto-server (sudo presto-server stop
seguido por sudo presto-server start).
```

- Spark
 - Backported [SPARK-22036: a BigDecimal multiplicação](#) às vezes retorna nula.

Problemas conhecidos

- MXNet não inclui bibliotecas OpenCV.
- O SparkR não está disponível para clusters criados usando um customizado porque o AMI R não é instalado por padrão nos nós do cluster.

Versão 5.11.3

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.11.3 da Amazon. As alterações são referentes à versão 5.11.2.

Data da versão inicial: 18 de julho de 2019

Alterações, melhorias e problemas resolvidos

- Atualizou o Amazon Linux padrão AMI para Amazon EMR para incluir atualizações importantes de segurança do kernel Linux, incluindo o problema de TCP SACK negação de serviço ([AWS-2019-005](#)).

Versão 5.11.2

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.11.2 da Amazon. As alterações são referentes à versão 5.11.1.

Data da versão inicial: 29 de agosto de 2018

Alterações, melhorias e problemas resolvidos

- Esta versão aborda uma possível vulnerabilidade de segurança.

Versão 5.11.1

As notas de lançamento a seguir incluem informações para a EMR versão 5.11.1 da Amazon. As alterações são relativas à EMR versão 5.11.0 da Amazon.

Data da versão inicial: 22 de janeiro de 2018

Alterações, melhorias e problemas resolvidos

- Atualizou o kernel Amazon Linux do default Amazon Linux AMI para Amazon para resolver vulnerabilidades associadas EMR à execução especulativa (CVE-2017-5715, -2017-5753 e -2017-5754). CVE CVE Para obter mais informações, consulte <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

Problemas conhecidos

- MXNet não inclui bibliotecas OpenCV.
- Por padrão, o Hive 2.3.2 define `hive.compute.query.using.stats=true`. Isso faz com que as consultas obtenham dados de estatísticas existentes em vez de diretamente dos dados, o que pode gerar confusão. Por exemplo, se você tiver uma tabela com `hive.compute.query.using.stats=true` e fizer upload de novos arquivos para a tabela LOCATION, a execução de uma consulta `SELECT COUNT(*)` na tabela retornará a contagem das estatísticas, e não selecionará as linhas adicionadas.

Como alternativa, use o comando `ANALYZE TABLE` para reunir novas estatísticas ou defina `hive.compute.query.using.stats=false`. Para obter mais informações, consulte [Estatísticas no Hive](#) na documentação do Apache Hive.

Versão 5.11.0

As notas de lançamento a seguir incluem informações para a EMR versão 5.11.0 da Amazon. As alterações são relativas à EMR versão 5.10.0 da Amazon.

Atualizações

Os aplicativos e os componentes a seguir foram atualizados nesta versão para incluir as seguintes versões.

- Hive 2.3.2
- Spark 2.2.1
- SDK para Java 1.11.238

Novos atributos

- Spark
 - Adicionada a configuração `spark.decommissioning.timeout.threshold`, que aprimora o comportamento de desativação do Spark ao usar instâncias spot. Para obter mais informações, consulte [Configurar o comportamento de desativação de nós](#).
 - [Foi adicionado o `aws-sagemaker-spark-sdk` componente ao Spark, que instala o Amazon SageMaker Spark e as dependências associadas para a integração do Spark com a Amazon SageMaker](#) Você pode usar o Amazon SageMaker Spark para criar pipelines de aprendizado de máquina (ML) do Spark usando os estágios da Amazon SageMaker Para obter mais informações, consulte o [readme do SageMaker Spark](#) sobre GitHub e como [usar o Apache Spark com a Amazon SageMaker no Amazon Developer](#) Guide. SageMaker

Problemas conhecidos

- MXNet não inclui bibliotecas OpenCV.
- Por padrão, o Hive 2.3.2 define `hive.compute.query.using.stats=true`. Isso faz com que as consultas obtenham dados de estatísticas existentes em vez de diretamente dos dados, o que pode gerar confusão. Por exemplo, se você tiver uma tabela com `hive.compute.query.using.stats=true` e fizer upload de novos arquivos para a tabela LOCATION, a execução de uma consulta `SELECT COUNT(*)` na tabela retornará a contagem das estatísticas, e não selecionará as linhas adicionadas.

Como alternativa, use o comando `ANALYZE TABLE` para reunir novas estatísticas ou defina `hive.compute.query.using.stats=false`. Para obter mais informações, consulte [Estatísticas no Hive](#) na documentação do Apache Hive.

Versão 5.10.0

As notas de lançamento a seguir incluem informações para a EMR versão 5.10.0 da Amazon. As alterações são relativas à EMR versão 5.9.0 da Amazon.

Atualizações

Os aplicativos e os componentes a seguir foram atualizados nesta versão para incluir as seguintes versões.

- AWS SDK for Java 1.11.221
- Hive 2.3.1
- Presto 0.187

Novos atributos

- Adicionado o suporte para autenticação do Kerberos. Para obter mais informações, consulte [Usar a autenticação Kerberos](#) no Amazon Management Guide EMR
- Foi adicionado suporte para IAM funções para EMRFS solicitações ao Amazon S3. Para obter mais informações, consulte [Configurar IAM funções para EMRFS solicitações ao Amazon S3](#) no Amazon EMR Management Guide.
- Foi adicionado suporte para GPU tipos de instância P2 e P3 baseados. Para obter mais informações, consulte Instâncias [Amazon EC2 P2](#) e Instâncias [Amazon EC2 P3](#). NVIDIAo driver 384.81 e o CUDA driver 9.0.176 são instalados nesses tipos de instância por padrão.
- O suporte adicionado para [Apache MXNet](#).

Alterações, melhorias e problemas resolvidos

- Presto
 - Foi adicionado suporte para usar o AWS Glue Data Catalog como metastore padrão do Hive. Para obter mais informações, consulte [Usando o Presto com o AWS Glue Data Catalog](#).
 - Suporte adicionado para [funções geoespaciais](#).
 - Suporte de [vazamento para disco](#) adicionado para uniões.
 - Suporte adicionado para o [Conector Redshift](#).
- Spark
 - Backported [SPARK-20640](#), que torna o tempo limite do rpc e as novas tentativas de valores de registro aleatório configuráveis, usando e propriedades. `spark.shuffle.registration.timeout` `spark.shuffle.registration.maxAttempts`

- Backported [SPARK-21549](#), que corrige um erro que ocorre ao gravar de forma personalizada OutputFormat em locais diferentes. HDFS
- Backport [Hadoop-13270](#)
- As bibliotecas Numpy, Scipy e Matplotlib foram removidas da Amazon base. EMR AMI Se forem necessárias para o aplicativo, essas bibliotecas estarão disponíveis no repositório do aplicativo. Portanto, você pode usar uma ação de bootstrap para instalá-las em todos os nós usando `yum install`.
- A EMR base da Amazon AMI não tem mais RPM pacotes de aplicativos incluídos, então os RPM pacotes não estão mais presentes nos nós do cluster. A Custom AMIs e a EMR base da Amazon AMI agora fazem referência ao repositório de RPM pacotes no Amazon S3.
- Devido à introdução do faturamento por segundo na AmazonEC2, o comportamento padrão de redução de escala agora é Encerrar na conclusão da tarefa, em vez de Encerrar na hora da instância. Para obter mais informações, consulte [Configurar redução da escala verticalmente do cluster](#).

Problemas conhecidos

- MXNet não inclui bibliotecas OpenCV.
- Por padrão, o Hive 2.3.1 define `hive.compute.query.using.stats=true`. Isso faz com que as consultas obtenham dados de estatísticas existentes em vez de diretamente dos dados, o que pode gerar confusão. Por exemplo, se você tiver uma tabela com `hive.compute.query.using.stats=true` e fizer upload de novos arquivos para a tabela LOCATION, a execução de uma consulta `SELECT COUNT(*)` na tabela retornará a contagem das estatísticas, e não selecionará as linhas adicionadas.

Como alternativa, use o comando `ANALYZE TABLE` para reunir novas estatísticas ou defina `hive.compute.query.using.stats=false`. Para obter mais informações, consulte [Estatísticas no Hive](#) na documentação do Apache Hive.

Versão 5.9.0

As notas de lançamento a seguir incluem informações para a EMR versão 5.9.0 da Amazon. As alterações são relativas à EMR versão 5.8.0 da Amazon.

Data do release: 5 de outubro de 2017

Última atualização de recursos: 12 de outubro de 2017

Atualizações

Os aplicativos e os componentes a seguir foram atualizados nesta versão para incluir as seguintes versões.

- AWS SDK for Java versão 1.11.183
- Flink 1.3.2
- Hue 4.0.1
- Pig 0.17.0
- Presto 0.184

Novos atributos

- Adição do suporte ao Livy (versão Livy 0.4.0 - em incubação). Para obter mais informações, consulte [Apache Livy](#).
- Adição de suporte para Hue Notebook para Spark.
- Foi adicionado suporte para EC2 instâncias Amazon da série i3 (12 de outubro de 2017).

Alterações, melhorias e problemas resolvidos

- Spark
 - Adição de um novo conjunto de recursos que ajudam a garantir que o Spark lide de uma forma mais fácil com o encerramento de nós devido a um redimensionamento manual ou uma solicitação de política de escalabilidade automática. Para obter mais informações, consulte [Configurar o comportamento de desativação de nós](#).
 - SSL é usado em vez de 3 DES para criptografia em trânsito para o serviço de transferência em bloco, o que melhora o desempenho ao usar tipos de EC2 instância da Amazon com AES -NI.
 - Retroportado para [SPARK-21494](#).
- Zeppelin
 - Portado para trás [ZEPPELIN-2377](#).
- HBase
 - Foi adicionado o patch [HBASE-18533](#), que permite valores adicionais para HBase BucketCache configuração usando a `hbase-site` classificação de configuração.

- Hue
 - Foi adicionado suporte ao AWS Glue Data Catalog para o editor de consultas Hive no Hue.
 - Por padrão, os superusuários no Hue podem acessar todos os arquivos que as EMR IAM funções da Amazon têm permissão para acessar. Os usuários recém-criados não têm automaticamente permissões para acessar o navegador de arquivos Amazon S3 e devem ter as permissões `filebrowser.s3_access` ativadas para o grupo deles.
- Resolveu um problema que fazia com que JSON os dados subjacentes criados usando o AWS Glue Data Catalog ficassem inacessíveis.

Problemas conhecidos

- A inicialização do cluster falha quando todos os aplicativos são instalados e o tamanho padrão do volume EBS raiz da Amazon não é alterado. Como solução alternativa, use o `aws emr create-cluster` comando do AWS CLI e especifique um `--ebs-root-volume-size` parâmetro maior.
- Por padrão, o Hive 2.3.0 define `hive.compute.query.using.stats=true`. Isso faz com que as consultas obtenham dados de estatísticas existentes em vez de diretamente dos dados, o que pode gerar confusão. Por exemplo, se você tiver uma tabela com `hive.compute.query.using.stats=true` e fizer upload de novos arquivos para a tabela LOCATION, a execução de uma consulta `SELECT COUNT(*)` na tabela retornará a contagem das estatísticas, e não selecionará as linhas adicionadas.

Como alternativa, use o comando `ANALYZE TABLE` para reunir novas estatísticas ou defina `hive.compute.query.using.stats=false`. Para obter mais informações, consulte [Estatísticas no Hive](#) na documentação do Apache Hive.

Versão 5.8.2

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.8.2 da Amazon. As alterações são referentes à versão 5.8.1.

Data da versão inicial: 29 de março de 2018

Alterações, melhorias e problemas resolvidos

- Atualizou o kernel Amazon Linux do default Amazon Linux AMI para Amazon EMR para resolver possíveis vulnerabilidades.

Versão 5.8.1

As notas de lançamento a seguir incluem informações para a EMR versão 5.8.1 da Amazon. As alterações são relativas à EMR versão 5.8.0 da Amazon.

Data da versão inicial: 22 de janeiro de 2018

Alterações, melhorias e problemas resolvidos

- Atualizou o kernel Amazon Linux do default Amazon Linux AMI para Amazon para resolver vulnerabilidades associadas EMR à execução especulativa (CVE-2017-5715, -2017-5753 e -2017-5754). CVE CVE Para obter mais informações, consulte <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

Versão 5.8.0

As notas de lançamento a seguir incluem informações para a EMR versão 5.8.0 da Amazon. As alterações são relativas à EMR versão 5.7.0 da Amazon.

Data da versão inicial: 10 de agosto de 2017

Última atualização de recurso: 25 de setembro de 2017

Atualizações

Os aplicativos e os componentes a seguir foram atualizados nesta versão para incluir as seguintes versões:

- AWS SDK 1.11.160
- Flink 1.3.1
- Hive 2.3.0. Para obter mais informações, consulte [Notas de versão](#) no site do Apache Hive.
- Spark 2.2.0. Para obter mais informações, consulte [Notas de versão](#) no site do Apache Spark.

Novos atributos

- Adição de suporte para visualização do histórico de aplicativos (25 de setembro de 2017). Para obter mais informações, consulte [Visualização do histórico de aplicativos](#) no Amazon EMR Management Guide.

Alterações, melhorias e problemas resolvidos

- Integração com o AWS Glue Data Catalog
 - Foi adicionada a capacidade do Hive e do Spark de usar SQL o AWS Glue Data Catalog como armazenamento de metadados do Hive. Para ter mais informações, consulte [Usando o AWS Glue Data Catalog como metastore para o Hive](#) e [Use o AWS Glue Data Catalog como metastore para o Spark SQL](#).
- Foi adicionado o histórico do aplicativo aos detalhes do cluster, o que permite visualizar dados históricos dos YARN aplicativos e detalhes adicionais dos aplicativos Spark. Para obter mais informações, consulte [Visualizar o histórico do aplicativo](#) no Amazon EMR Management Guide.
- Oozie
 - Porta traseira [OOZIE-2748](#).
- Hue
 - [Porta traseira - 5859 HUE](#)
- HBase
 - Patch adicionado para expor a hora de início do servidor HBase mestre por meio de Java Management Extensions (JMX) usando `getMasterInitializedTime`.
 - Adicionado um patch que melhora a hora de início do cluster.

Problemas conhecidos

- A inicialização do cluster falha quando todos os aplicativos são instalados e o tamanho padrão do volume EBS raiz da Amazon não é alterado. Como solução alternativa, use o `aws emr create-cluster` comando do AWS CLI e especifique um `--efs-root-volume-size` parâmetro maior.
- Por padrão, o Hive 2.3.0 define `hive.compute.query.using.stats=true`. Isso faz com que as consultas obtenham dados de estatísticas existentes em vez de diretamente dos dados, o que pode gerar confusão. Por exemplo, se você tiver uma tabela com `hive.compute.query.using.stats=true` e fizer upload de novos arquivos para a tabela LOCATION, a execução de uma consulta `SELECT COUNT(*)` na tabela retornará a contagem das estatísticas, e não selecionará as linhas adicionadas.

Como alternativa, use o comando `ANALYZE TABLE` para reunir novas estatísticas ou defina `hive.compute.query.using.stats=false`. Para obter mais informações, consulte [Estatísticas no Hive](#) na documentação do Apache Hive.

- Spark: ao usar o Spark, há um problema de vazamento no manipulador de arquivos com o daemon apppusher, o que pode ocorrer para um trabalho de execução prolongada do Spark depois de várias horas ou dias. Para corrigir o problema, conecte-se ao nó principal e digite `sudo /etc/init.d/apppusher stop`. Isso interrompe o daemon apppusher, que a Amazon EMR reiniciará automaticamente.
- Application history
 - Os dados históricos dos executores inativos do Spark não está disponível.
 - O histórico do aplicativo não está disponível para clusters que usam uma configuração de segurança para habilitar a criptografia em andamento.

Versão 5.7.0

As notas de lançamento a seguir incluem informações para a versão EMR 5.7.0 da Amazon. As alterações são relativas à EMR versão 5.6.0 da Amazon.

Data do release: 13 de julho de 2017

Atualizações

- Flink 1.3.0
- Phoenix 4.11.0
- Zeppelin 0.7.2

Novos atributos

- Foi adicionada a capacidade de especificar um Amazon Linux personalizado AMI ao criar um cluster. Para obter mais informações, consulte [Usando um personalizado AMI](#).

Alterações, melhorias e problemas resolvidos

- HBase
 - Capacidade adicional para configurar clusters de HBase réplica de leitura. Consulte [Usar um cluster de réplica de leitura](#).
 - Várias correções de erros e melhorias
- Presto: adicionada a capacidade de configurar `node.properties`.

- YARN- capacidade adicional de configurar `container-log4j.properties`
- Sqoop - backported [SQOOP-2880](#), que introduz um argumento que permite definir o diretório temporário do Sqoop.

Versão 5.6.0

As notas de lançamento a seguir incluem informações para a versão EMR 5.6.0 da Amazon. As alterações são relativas à versão EMR 5.5.0 da Amazon.

Data do release: 5 de junho de 2017

Atualizações

- Flink 1.2.1
- HBase1.3.1
- Mahout 0.13.0. Essa é a primeira versão do Mahout a oferecer suporte ao Spark 2.x na Amazon EMR versão 5.0 e posterior.
- Spark 2.1.1

Alterações, melhorias e problemas resolvidos

- Presto
 - Foi adicionada a capacidade de SSL TLS habilitar/proteger a comunicação entre os nós do Presto ativando a criptografia em trânsito usando uma configuração de segurança. Para obter mais informações, consulte [Criptografia de dados em trânsito](#).
 - [Presto 7661](#) enviado para backport, adiciona a opção VERBOSE à instrução EXPLAIN ANALYZE para relatar estatísticas de baixo nível mais detalhadas sobre um plano de consulta.

Versão 5.5.3

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.5.3 da Amazon. As alterações são referentes à versão 5.5.2.

Data da versão inicial: 29 de agosto de 2018

Alterações, melhorias e problemas resolvidos

- Esta versão aborda uma possível vulnerabilidade de segurança.

Versão 5.5.2

As notas de lançamento a seguir incluem informações sobre a EMR versão 5.5.2 da Amazon. As alterações são referentes à versão 5.5.1.

Data da versão inicial: 29 de março de 2018

Alterações, melhorias e problemas resolvidos

- Atualizou o kernel Amazon Linux do default Amazon Linux AMI para Amazon EMR para resolver possíveis vulnerabilidades.

Versão 5.5.1

As notas de lançamento a seguir incluem informações para a versão EMR 5.5.1 da Amazon. As alterações são relativas à versão EMR 5.5.0 da Amazon.

Data da versão inicial: 22 de janeiro de 2018

Alterações, melhorias e problemas resolvidos

- Atualizou o kernel Amazon Linux do default Amazon Linux AMI para Amazon para resolver vulnerabilidades associadas EMR à execução especulativa (CVE-2017-5715, -2017-5753 e -2017-5754). CVE CVE Para obter mais informações, consulte <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

Versão 5.5.0

As notas de lançamento a seguir incluem informações para a versão EMR 5.5.0 da Amazon. As alterações são relativas à versão EMR 5.4.0 da Amazon.

Data do release: 26 de abril de 2017

Atualizações

- Hue 3.12

- Presto 0.170
- Zeppelin 0.7.1
- ZooKeeper 3.4.10

Alterações, melhorias e problemas resolvidos

- Spark
 - [Correção do Backported Spark Patch \(SPARK-20115\) DAGScheduler para recalculer todos os blocos de reprodução aleatória perdidos quando o serviço externo de reprodução aleatória não está disponível na versão 2.1.0 do Spark, incluída nesta versão.](#)
- Flink
 - O Flink agora é compilado com o Scala 2.11. Se você usa o Scala API e as bibliotecas, recomendamos usar o Scala 2.11 em seus projetos.
 - Tratado um problema em que os padrões HADOOP_CONF_DIR e YARN_CONF_DIR não estavam definidos corretamente, portanto havia falha no funcionamento de `start-scala-shell.sh`. Também foi adicionada a possibilidade de definir esses valores usando `env.hadoop.conf.dir` e `env.yarn.conf.dir` em `/etc/flink/conf/flink-conf.yaml` ou na classificação de configuração `flink-conf`.
 - Introduziu um novo comando EMR específico, `flink-scala-shell` como invólucro para `start-scala-shell.sh`. Recomendamos o uso desse comando, em vez de `start-scala-shell`. O novo comando simplifica a execução. Por exemplo, `flink-scala-shell -n 2` inicia um shell Scala Flink com um paralelismo de tarefa de 2.
 - Introduziu um novo comando EMR específico, `flink-yarn-session` como invólucro para `yarn-session.sh`. Recomendamos o uso desse comando, em vez de `yarn-session`. O novo comando simplifica a execução. Por exemplo, `flink-yarn-session -d -n 2` inicia uma sessão de longa execução do Flink em um estado desanexado com dois gerenciadores de tarefas.
 - O [commons httpclient endereçado \(FLINK-6125\) não está mais sombreado](#) no Flink 1.2.
- Presto
 - Foi adicionado suporte para LDAP autenticação. O uso LDAP com o Presto na Amazon EMR exige que você habilite o HTTPS acesso para o coordenador do Presto (`http-server.https.enabled=trueemconfig.properties`). Para obter detalhes de configuração, consulte a [LDAPautenticação](#) na documentação do Presto.
 - O suporte adicionado para `SHOW GRANTS`.

- Amazon EMR Base Linux AMI
 - Os EMR lançamentos da Amazon agora são baseados no Amazon Linux 2017.03. Para obter mais informações, consulte as notas de [versão do Amazon Linux AMI 2017.03](#).
 - O Python 2.6 foi removido da imagem Linux EMR básica da Amazon. Python 2.7 e 3.4 estão instalados por padrão. Você pode instalar o Python 2.6 manualmente, se necessário.

Versão 5.4.0

As notas de lançamento a seguir incluem informações para a versão EMR 5.4.0 da Amazon. As alterações são relativas à versão EMR 5.3.0 da Amazon.

Data do release: 08 de março de 2017

Atualizações

As seguintes atualizações estão disponíveis nesta versão:

- Atualizado para Flink 1.2.0
- Atualizado para Hbase 1.3.0
- Atualizado para Phoenix 4.9.0

Note

Se você atualizar de uma versão anterior da Amazon EMR para a EMR versão 5.4.0 ou posterior da Amazon e usar indexação secundária, atualize os índices locais conforme descrito na documentação do [Apache](#) Phoenix. A Amazon EMR remove as configurações necessárias da `hbase-site` classificação, mas os índices precisam ser preenchidos novamente. O sistema oferece suporte a atualizações de índices online e offline. As atualizações online são o padrão, o que significa que os índices são preenchidos novamente durante a inicialização de clientes do Phoenix versão 4.8.0 ou posterior. Para especificar atualizações off-line, defina a `phoenix.client.localIndexUpgrade` configuração como falsa na `phoenix-site` classificação e, em seguida, no nó principal SSH a ser executado `psql [zookeeper] -1`.

- Atualizado para Presto 0.166
- Atualizado para Zeppelin 0.7.0

Alterações e melhorias

A seguir estão as alterações feitas nos EMR lançamentos da Amazon para a etiqueta de lançamento emr-5.4.0:

- Adicionado o suporte para instâncias r4. Veja os [tipos de EC2 instância da Amazon](#).

Versão 5.3.1

As notas de lançamento a seguir incluem informações para a versão EMR 5.3.1 da Amazon. As alterações são relativas à versão EMR 5.3.0 da Amazon.

Data da versão: 7 de fevereiro de 2017

Pequenas alterações nos patches do Zeppelin e na atualização do padrão para a Amazon. AMI EMR

Versão 5.3.0

As notas de lançamento a seguir incluem informações para a versão EMR 5.3.0 da Amazon. As alterações são relativas à versão EMR 5.2.1 da Amazon.

Data do release: 26 de janeiro de 2017

Atualizações

As seguintes atualizações estão disponíveis nesta versão:

- Atualizado para Hive 2.1.1
- Atualizado para Hue 3.11.0
- Atualizado para Spark 2.1.0
- Atualizado para Oozie 4.3.0
- Atualizado para Flink 1.1.4

Alterações e melhorias

A seguir estão as alterações feitas nos EMR lançamentos da Amazon para a etiqueta de lançamento emr-5.3.0:

- Adicionado um patch para o Hue que permite usar a configuração `interpreters_shown_on_wheel` para definir o que intérpretes mostram primeiro na roda de seleção do bloco de anotações, independentemente de sua ordem no arquivo `hue.ini`.
- Adicionada a classificação de configuração `hive-parquet-logging`, que pode ser usada para configurar os valores no arquivo `parquet-logging.properties` do Hive.

Versão 5.2.2

As notas de lançamento a seguir incluem informações para a versão EMR 5.2.2 da Amazon. As alterações são relativas à versão EMR 5.2.1 da Amazon.

Data do release: 2 de maio de 2017

Problemas conhecidos das versões anteriores que foram resolvidos

- Backported [SPARK-194459](#), que resolve um problema em que a leitura de uma ORC tabela com colunas `char`/`varchar` pode falhar.

Versão 5.2.1

As notas de lançamento a seguir incluem informações para a versão EMR 5.2.1 da Amazon. As alterações são relativas à versão EMR 5.2.0 da Amazon.

Data do release: 29 de dezembro de 2016

Atualizações

As seguintes atualizações estão disponíveis nesta versão:

- Atualizado para Presto 0.157.1. Para obter mais informações, consulte [Notas de versão do Presto](#) na documentação do Presto.
- Zookeeper atualizado para 3.4.9. Para obter mais informações, consulte as [notas de ZooKeeper lançamento](#) na ZooKeeper documentação do Apache.

Alterações e melhorias

A seguir estão as alterações feitas nos EMR lançamentos da Amazon para a etiqueta de lançamento `emr-5.2.1`:

- Foi adicionado suporte para o tipo de EC2 instância Amazon m4.16xlarge na Amazon EMR versão 4.8.3 e posterior, excluindo 5.0.0, 5.0.3 e 5.2.0.
- Os EMR lançamentos da Amazon agora são baseados no Amazon Linux 2016.09. Para obter mais informações, consulte <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.
- A localização do Flink e os caminhos de YARN configuração agora são definidos por padrão, pois `/etc/default/flink` você não precisa definir as variáveis de ambiente `FLINK_CONF_DIR` e `HADOOP_CONF_DIR` ao executar os scripts `flink` ou os scripts do `yarn-session.sh` driver, para iniciar trabalhos do Flink.
- Foi adicionado suporte para a `FlinkKinesisConsumer` aula.

Problemas conhecidos das versões anteriores que foram resolvidos

- Corrigido um problema no Hadoop em que o `ReplicationMonitor` thread podia ficar preso por um longo tempo devido a uma corrida entre a replicação e a exclusão do mesmo arquivo em um grande cluster.
- Corrigido um problema em que `ControlledJob # toString` falhava com uma exceção de ponteiro nulo (NPE) quando o status do trabalho não era atualizado com êxito.

Versão 5.2.0

As notas de lançamento a seguir incluem informações para a versão EMR 5.2.0 da Amazon. As alterações são relativas à versão EMR 5.1.0 da Amazon.

Data do release: 21 de novembro de 2016

Alterações e melhorias

As seguintes alterações e melhorias estão disponíveis nesta versão:

- Adicionado o modo de armazenamento Amazon S3 para HBase
- Permite que você especifique uma localização do Amazon S3 para o HBase rootdir. Para obter mais informações, consulte [HBase no Amazon S3](#).

Atualizações

As seguintes atualizações estão disponíveis nesta versão:

- Atualizado para Spark 2.0.2

Problemas conhecidos das versões anteriores que foram resolvidos

- Corrigido um problema com /mnt sendo restrito a 2 TB em tipos EBS de instância somente.
- Corrigido um problema com o controlador de instância e com os logs do logpusher serem a saída para seus arquivos .out correspondentes, em vez de para seus arquivos normais .log configurados com log4j, que mudam de hora em hora. Os arquivos .out não mudam, portanto isso eventualmente encheria a partição /emr. Esse problema afeta apenas os tipos de instância de máquina virtual (HVM) de hardware.

Versão 5.1.0

As notas de lançamento a seguir incluem informações para a versão EMR 5.1.0 da Amazon. As alterações são relativas à versão EMR 5.0.0 da Amazon.

Data do release: 03 de novembro de 2016

Alterações e melhorias

As seguintes alterações e melhorias estão disponíveis nesta versão:

- Adicionado o suporte para Flink 1.1.3.
- O Presto foi adicionado como uma opção na seção bloco de anotações do Hue.

Atualizações

As seguintes atualizações estão disponíveis nesta versão:

- Atualizado para 1.2.3 HBase
- Atualizado para Zeppelin 0.6.2

Problemas conhecidos das versões anteriores que foram resolvidos

- Corrigido um problema com as consultas Tez no Amazon S3 ORC com arquivos que não funcionavam tão bem quanto nas versões anteriores do Amazon EMR 4.x.

Versão 5.0.3

As notas de lançamento a seguir incluem informações para a versão EMR 5.0.3 da Amazon. As alterações são relativas à versão EMR 5.0.0 da Amazon.

Data do release: 24 de outubro de 2016

Atualizações

As seguintes atualizações estão disponíveis nesta versão:

- Atualizado para Hadoop 2.7.3
- Atualizado para Presto 0.152.3, que inclui o suporte para a interface da web do Presto. Você pode acessar a interface da web do Presto no coordenador do Presto usando a porta 8889. Para obter mais informações sobre a interface da Web do Presto, consulte [Interface da Web](#) na documentação do Presto.
- Atualizado para Spark 2.0.1
- Os EMR lançamentos da Amazon agora são baseados no Amazon Linux 2016.09. Para obter mais informações, consulte <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.

Versão 5.0.0

Data do release: 27 de julho de 2016

Atualizações

As seguintes atualizações estão disponíveis nesta versão:

- Atualizado para Hive 2.1
- Atualizado para Presto 0.150
- Atualizado para Spark 2.0
- Atualizado para Hue 3.10.0
- Atualizado para Pig 0.16.0
- Atualizado para Tez 0.8.4
- Atualizado para Zeppelin 0.6.1

Alterações e melhorias

A seguir estão as alterações feitas nos EMR lançamentos da Amazon para a etiqueta de lançamento emr-5.0.0 ou superior:

- A Amazon EMR oferece suporte às versões mais recentes de código aberto do Hive (versão 2.1) e do Pig (versão 0.16.0). Se você já usou o Hive ou o Pig EMR na Amazon no passado, isso pode afetar alguns casos de uso. Para obter mais informações, consulte [Hive](#) e [Pig](#).
- O mecanismo de execução padrão para o Hive e o Pig agora é o Tez. Para alterar isso, você deve editar os valores apropriados nas classificações de configuração `hive-site` e `pig-properties`, respectivamente.
- Um recurso de etapa aprimorada de depuração foi adicionado, o que permite que você veja a causa raiz de falhas de etapa se o serviço puder determinar a causa. Para obter mais informações, consulte [Depuração de etapas aprimorada](#) no Amazon EMR Management Guide.
- Os aplicativos que, anteriormente, terminavam com "-Sandbox" não têm mais esse sufixo. Isso pode inutilizar sua automação, por exemplo, se você estiver usando scripts para iniciar clusters com esses aplicativos. A tabela a seguir mostra os nomes dos aplicativos no Amazon EMR 4.7.2 versus no Amazon EMR 5.0.0.

Alterações dos nomes de aplicativos

Amazon EMR 4.7.2	Amazon EMR 5.0.0
Oozie-Sandbox	Oozie
Presto-Sandbox	Presto
Sqoop-Sandbox	Sqoop
Zeppelin-Sandbox	Zeppelin
ZooKeeper-Caixa de areia	ZooKeeper

- O Spark agora está compilado para Scala 2.11.
- O Java 8 agora é o padrão JVM. Todas as aplicações são executadas usando o runtime do Java 8. Não há alterações em qualquer destino de código de bytes da aplicação. A maioria dos aplicativos continuam a usar o Java 7 como destino.
- O Zeppelin agora inclui recursos de autenticação. Para obter mais informações, consulte [Zeppelin](#).

- Adicionado o suporte para configurações de segurança, que permitem criar e aplicar opções de criptografia com mais facilidade. Para obter mais informações, consulte [Criptografia de dados](#).

Versão 4.9.5

As notas de lançamento a seguir incluem informações sobre a EMR versão 4.9.5 da Amazon. As alterações são referentes à versão 4.9.4.

Data da versão inicial: 29 de agosto de 2018

Alterações, melhorias e problemas resolvidos

- HBase
 - Esta versão aborda uma possível vulnerabilidade de segurança.

Versão 4.9.4

As notas de lançamento a seguir incluem informações sobre a EMR versão 4.9.4 da Amazon. As alterações são referentes à versão 4.9.3.

Data da versão inicial: 29 de março de 2018

Alterações, melhorias e problemas resolvidos

- Atualizou o kernel Amazon Linux do defaultAmazon Linux AMI para Amazon EMR para resolver possíveis vulnerabilidades.

Versão 4.9.3

As notas de lançamento a seguir incluem informações para a versão EMR 4.9.3 da Amazon. As alterações são relativas à versão EMR 4.9.2 da Amazon.

Data da versão inicial: 22 de janeiro de 2018

Alterações, melhorias e problemas resolvidos

- Atualizou o kernel Amazon Linux do defaultAmazon Linux AMI para Amazon para resolver vulnerabilidades associadas EMR à execução especulativa (CVE-2017-5715, -2017-5753 e

-2017-5754). CVE CVE Para obter mais informações, consulte <https://aws.amazon.com/security/security-bulletins/AWS-2018-013/>.

Versão 4.9.2

As notas de lançamento a seguir incluem informações para a versão EMR 4.9.2 da Amazon. As alterações são relativas à versão EMR 4.9.1 da Amazon.

Data do release: 13 de julho de 2017

Pequenas alterações, correções de erros e melhorias foram feitas nesta versão.

Versão 4.9.1

As notas de lançamento a seguir incluem informações para a versão EMR 4.9.1 da Amazon. As alterações são relativas à versão EMR 4.8.4 da Amazon.

Data do release: 10 de abril de 2017

Problemas conhecidos das versões anteriores que foram resolvidos

- [Portas traseiras de HIVE-9976 e -10106 HIVE](#)
- Corrigido um problema em YARN que um grande número de nós (mais de 2.000) e contêineres (mais de 5.000) causava um erro de falta de memória, por exemplo: "Exception in thread 'main' java.lang.OutOfMemoryError".

Alterações e melhorias

A seguir estão as alterações feitas nos EMR lançamentos da Amazon para a etiqueta de lançamento emr-4.9.1:

- Os EMR lançamentos da Amazon agora são baseados no Amazon Linux 2017.03. Para obter mais informações, consulte <https://aws.amazon.com/amazon-linux-ami/2017.03-release-notes/>.
- O Python 2.6 foi removido da imagem Linux EMR básica da Amazon. Você pode instalar o Python 2.6 manualmente, se necessário.

Versão 4.8.4

As notas de lançamento a seguir incluem informações para a versão EMR 4.8.4 da Amazon. As alterações são relativas à versão EMR 4.8.3 da Amazon.

Data do release: 7 de fevereiro de 2017

Pequenas alterações, correções de erros e melhorias foram feitas nesta versão.

Versão 4.8.3

As notas de lançamento a seguir incluem informações para a versão EMR 4.8.3 da Amazon. As alterações são relativas à versão EMR 4.8.2 da Amazon.

Data do release: 29 de dezembro de 2016

Atualizações

As seguintes atualizações estão disponíveis nesta versão:

- Atualizado para Presto 0.157.1. Para obter mais informações, consulte [Notas de versão do Presto](#) na documentação do Presto.
- Atualizado para Spark 1.6.3. Para obter mais informações, consulte [Spark release notes](#) na documentação do Apache Spark.
- Atualizado para ZooKeeper 3.4.9. Para obter mais informações, consulte as [notas de ZooKeeper lançamento](#) na ZooKeeper documentação do Apache.

Alterações e melhorias

A seguir estão as alterações feitas nos EMR lançamentos da Amazon para a etiqueta de lançamento emr-4.8.3:

- Foi adicionado suporte para o tipo de EC2 instância Amazon m4.16xlarge na Amazon EMR versão 4.8.3 e posterior, excluindo 5.0.0, 5.0.3 e 5.2.0.
- Os EMR lançamentos da Amazon agora são baseados no Amazon Linux 2016.09. Para obter mais informações, consulte <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.

Problemas conhecidos das versões anteriores que foram resolvidos

- Corrigido um problema no Hadoop em que o ReplicationMonitor thread podia ficar preso por um longo tempo devido a uma corrida entre a replicação e a exclusão do mesmo arquivo em um grande cluster.
- Corrigido um problema em que ControlledJob # toString falhava com uma exceção de ponteiro nulo (NPE) quando o status do trabalho não era atualizado com êxito.

Versão 4.8.2

As notas de lançamento a seguir incluem informações para a versão EMR 4.8.2 da Amazon. As alterações são relativas à versão EMR 4.8.0 da Amazon.

Data do release: 24 de outubro de 2016

Atualizações

As seguintes atualizações estão disponíveis nesta versão:

- Atualizado para Hadoop 2.7.3
- Atualizado para Presto 0.152.3, que inclui o suporte para a interface da web do Presto. Você pode acessar a interface da web do Presto no coordenador do Presto usando a porta 8889. Para obter mais informações sobre a interface da Web do Presto, consulte [Interface da Web](#) na documentação do Presto.
- Os EMR lançamentos da Amazon agora são baseados no Amazon Linux 2016.09. Para obter mais informações, consulte <https://aws.amazon.com/amazon-linux-ami/2016.09-release-notes/>.

Versão 4.8.0

Data do release: 7 de setembro de 2016

Atualizações

As seguintes atualizações estão disponíveis nesta versão:

- Atualizado para 1.2.2 HBase
- Atualizado para Presto-Sandbox 0.151
- Atualizado para Tez 0.8.4

- Atualizado para Zeppelin-Sandbox 0.6.1

Alterações e melhorias

A seguir estão as alterações feitas nos EMR lançamentos da Amazon para a etiqueta de lançamento emr-4.8.0:

- Corrigido um problema YARN em que eles ApplicationMaster tentavam limpar contêineres que não existem mais porque suas instâncias foram encerradas.
- Foram corrigidas as ações hive-server2 URL para Hive2 nos exemplos do Oozie.
- Adicionado o suporte para catálogos Presto adicionais.
- [Patches retroportados: HIVE-8948, HIVE -12679, -13405, -3116, -12689 HIVE PHOENIX HADOOP](#)
- Adicionado o suporte para configurações de segurança, que permitem criar e aplicar opções de criptografia com mais facilidade. Para obter mais informações, consulte [Criptografia de dados](#).

Versão 4.7.2

As notas de lançamento a seguir incluem informações sobre o Amazon EMR 4.7.2.

Data do release: 15 de julho de 2016

Recursos

Os seguintes recursos estão disponíveis nesta versão:

- Atualizado para Mahout 0.12.2
- Atualizado para Presto 0.148
- Atualizado para Spark 1.6.2
- Agora você pode criar um AWSCredentialsProvider para uso com EMRFS o uso de um URI como parâmetro. Para obter mais informações, consulte [Criar um AWSCredentialsProvider formulário EMRFS](#).
- EMRFS agora permite que os usuários configurem um endpoint personalizado do DynamoDB para seus metadados do Consistent View usando a propriedade em `fs.s3.consistent.dynamodb.endpoint emrfs-site.xml`
- Adicionado um script em `/usr/bin` chamado `spark-example`, que encapsula `/usr/lib/spark/spark/bin/run-example` para que você possa executar exemplos diretamente. Por

exemplo, para executar o SparkPi exemplo que vem com a distribuição do Spark, você pode executar a `spark-example SparkPi 100` partir da linha de comando ou usando `command-runner.jar` como uma etapa noAPI.

Problemas conhecidos das versões anteriores que foram resolvidos

- Corrigido um problema em que o Oozie não tinha o `spark-assembly.jar` no local correto quando o Spark também estava instalado, o que resultava em falha para iniciar aplicativos do Spark com o Oozie.
- Corrigido um problema com o registro baseado no Spark Log4J em contêineres. YARN

Versão 4.7.1

Data do release: 10 de junho de 2016

Problemas conhecidos das versões anteriores que foram resolvidos

- Corrigido um problema que estendia o tempo de inicialização dos clusters lançados em um VPC com sub-redes privadas. O bug afetou apenas os clusters lançados com a EMR versão 4.7.0 da Amazon.
- Corrigido um problema que manipulava indevidamente a listagem de arquivos na Amazon EMR para clusters lançados com a versão EMR 4.7.0 da Amazon.

Versão 4.7.0

Important

O Amazon EMR 4.7.0 está obsoleto. Em vez disso, use o Amazon EMR 4.7.1 ou posterior.

Data do release: 2 de junho de 2016

Recursos

Os seguintes recursos estão disponíveis nesta versão:

- Adicionado o Apache Phoenix 4.7.0

- Adicionado o Apache Tez 0.8.3
- Atualizado para 1.2.1 HBase
- Atualizado para Mahout 0.12.0
- Atualizado para Presto 0.147
- Atualizou o AWS SDK for Java para 1.10.75
- O sinalizador final foi removido da propriedade `mapreduce.cluster.local.dir` em `mapred-site.xml` para permitir que os usuários executem o Pig no modo local.

JDBCDrivers do Amazon Redshift disponíveis no cluster

JDBCOs drivers do Amazon Redshift agora estão incluídos em `/usr/share/aws/redshift/jdbc` `/usr/share/aws/redshift/jdbc/RedshiftJDBC41.jar` é o driver do Amazon Redshift JDBC compatível com 4.1 `/usr/share/aws/redshift/jdbc/RedshiftJDBC4.jar` e é o driver do Amazon Redshift JDBC compatível com 4.0. Para obter mais informações, consulte [Configurar uma JDBC conexão](#) no Guia de gerenciamento do Amazon Redshift.

Java 8

Com exceção do Presto, o Open JDK 1.7 é o padrão JDK usado para todos os aplicativos. No entanto, o Open JDK 1.7 e o 1.8 estão instalados. Para obter mais informações sobre como configurar `JAVA_HOME` para aplicações, consulte [Configurar aplicações para usar Java 8](#).

Problemas conhecidos das versões anteriores que foram resolvidos

- Corrigido um problema de kernel que afetou significativamente o desempenho em volumes otimizados de taxa de transferência HDD (st1) da Amazon EMR no EBS `emr-4.6.0`.
- Corrigido um problema em que um cluster falharia se alguma zona de HDFS criptografia fosse especificada sem escolher o Hadoop como aplicativo.
- Alterou a política de HDFS gravação padrão de `RoundRobin` para `AvailableSpaceVolumeChoosingPolicy`. Alguns volumes não foram utilizados adequadamente com a `RoundRobin` configuração, o que resultou em falhas nos nós principais e na falta HDFS de confiabilidade.
- Corrigido um problema com o `EMRFSCLI`, que causaria uma exceção ao criar a tabela de metadados padrão do DynamoDB para visualizações consistentes.
- Corrigido um problema de impasse `EMRFS` que potencialmente ocorria durante operações de renomeação e cópia de várias partes.

- Foi corrigido um problema EMRFS que fazia com que o CopyPart tamanho padrão fosse 5 MB. O padrão agora está definido corretamente como 128 MB.
- Corrigido um problema com a configuração de inicialização do Zeppelin que potencialmente impedia a interrupção do serviço.
- Corrigido um problema com o Spark e o Zeppelin, que impedia você de usar o `s3a://` URI esquema porque ele não `/usr/lib/hadoop/hadoop-aws.jar` estava carregado corretamente em seus respectivos classpath.
- Portado para trás [HUE-2484](#).
- Reportou um [commit](#) do Hue 3.9.0 (não JIRA existe) para corrigir um problema com a amostra do navegador. HBase
- Portado para trás [HIVE-9073](#).

Versão 4.6.0

Data do release: 21 de abril de 2016

Recursos

Os seguintes recursos estão disponíveis nesta versão:

- Adicionado HBase 1.2.0
- Adicionado o ZooKeeper-Sandbox 3.4.8
- Atualizado para Presto-Sandbox 0.143
- Os EMR lançamentos da Amazon agora são baseados no Amazon Linux 2016.03.0. Para obter mais informações, consulte <https://aws.amazon.com/amazon-linux-ami/2016.03-release-notes/>.

Problema que afeta os tipos de volume com taxa de transferência otimizada HDD (st1) EBS

Um problema nas versões 4.2 e superiores do kernel Linux afeta significativamente o desempenho nos volumes Throughput Optimized HDD (st1) EBS do. EMR Esta versão (emr-4.6.0) usa uma versão do kernel 4.4.5 e, portanto, é afetada. Portanto, recomendamos não usar o emr-4.6.0 se você quiser usar volumes st1. EBS Você pode usar o emr-4.5.0 ou EMR versões anteriores da Amazon com st1 sem impacto. Além disso, fornecemos a correção com futuras versões.

Padrões do Python

O Python 3.4 agora está instalado por padrão, mas o Python 2.7 permanece como o sistema padrão. Você pode configurar o Python 3.4 como o padrão do sistema usando uma ação de bootstrap; você pode usar a configuração API para definir `PYSPARK_PYTHON` export to `/usr/bin/python3.4` na `spark-env` classificação para afetar a versão do Python usada por PySpark

Java 8

Com exceção do Presto, o Open JDK 1.7 é o padrão JDK usado para todos os aplicativos. No entanto, o Open JDK 1.7 e o 1.8 estão instalados. Para obter mais informações sobre como configurar `JAVA_HOME` para aplicações, consulte [Configurar aplicações para usar Java 8](#).

Problemas conhecidos das versões anteriores que foram resolvidos

- Corrigido um problema em que, às vezes, ocorria uma falha aleatória no provisionamento de aplicativos devido a uma senha gerada.
- Anteriormente, `mysql` estava instalado em todos os nós. Agora, ele só está instalado na instância principal e somente se o aplicativo escolhido incluir `mysql-server` como componente. Atualmente, os seguintes aplicativos incluem o `mysql-server` componente: HiveHCatalog, Hue, Presto-Sandbox e Sqoop-Sandbox.
- Alterado do padrão `32 yarn.scheduler.maximum-allocation-vcores` para 80, o que corrige um problema introduzido no `emr-4.4.0` que ocorre principalmente com o Spark ao usar a `maximizeResourceAllocation` opção em um cluster cujo tipo de instância principal é um dos poucos tipos de instância grandes que têm YARN vcores definidos acima de 32; ou seja, `c4.8xlarge`, `cc2.8xlarge`, `hs1.8xlarge`, `i2.8xlarge`, `m2.4xlarge`, `r3.8xlarge`, `d2.8xlarge` grande ou `m4.10xlarge` foram afetados por esse problema.
- O `s3-dist-cp` agora é usado para todas as indicações ao EMRFS Amazon S3 e não é mais usado em um diretório temporário. HDFS
- Corrigido um problema com o tratamento de exceções para os multipart uploads de criptografia no lado do cliente.
- Adicionada uma opção para permitir que os usuários alterem a classe de armazenamento do Amazon S3. Por padrão, essa configuração é `STANDARD`. A configuração da classificação de configuração `emrfs-site` é `fs.s3.storageClass` e os valores possíveis são `STANDARD`, `STANDARD_IA` e `REDUCED_REDUNDANCY`. Para obter mais informações sobre classes de armazenamento, consulte [Classes de armazenamento](#) no Guia do usuário do Amazon Simple Storage Service.

Versão 4.5.0

Data do release: 4 de abril de 2016

Recursos

Os seguintes recursos estão disponíveis nesta versão:

- Atualizado para Spark 1.6.1
- Atualizado para Hadoop 2.7.2
- Atualizado para Presto 0.140
- Foi adicionado AWS KMS suporte para criptografia do lado do servidor Amazon S3.

Problemas conhecidos das versões anteriores que foram resolvidos

- Corrigido um problema em que os servidores My SQL e Apache não iniciavam após a reinicialização de um nó.
- Corrigido um problema em IMPORT que não funcionava corretamente com tabelas não particionadas armazenadas no Amazon S3
- Corrigido um problema em que o Presto exigia que o diretório de preparo fosse /mnt/tmp em vez de /tmp ao gravar em tabelas do Hive.

Versão 4.4.0

Data do release: 14 de março de 2016

Recursos

Os seguintes recursos estão disponíveis nesta versão:

- Adicionado HCatalog 1.0.0
- Adicionado o Sqoop-Sandbox 1.4.6
- Atualizado para Presto 0.136
- Atualizado para Zeppelin 0.5.6
- Atualizado para Mahout 0.11.1
- Habilitada a `dynamicResourceAllocation` por padrão.

- Adicionada uma tabela de todas as classificações de configuração para a versão. Para obter mais informações, consulte a tabela de classificações de configuração em [Configurar aplicações](#).

Problemas conhecidos das versões anteriores que foram resolvidos

- Corrigido um problema em que a `maximizeResourceAllocation` configuração não reservava memória suficiente para YARN ApplicationMaster daemons.
- Corrigido um problema encontrado com um personalizadoDNS. Se alguma entrada em `resolve.conf` preceder as entradas personalizadas fornecidas, as entradas personalizadas não serão resolvíveis. Esse comportamento foi afetado por clusters em um em VPC que o servidor de VPC nomes padrão é inserido como a entrada superior em `resolve.conf`.
- Corrigido um problema em que o Python padrão mudou para a versão 2.7 e boto não estava instalado para essa versão.
- Corrigido um problema em que YARN contêineres e aplicativos Spark geravam um arquivo de banco de dados Round Robin (rrd) exclusivo do Ganglia, o que resultava no preenchimento do primeiro disco conectado à instância. Por causa dessa correção, as métricas em nível de YARN contêiner foram desativadas e as métricas em nível de aplicativo do Spark foram desativadas.
- Corrigido um problema no log pusher em que ele excluía todas as pastas de log vazias. O efeito foi que o Hive não CLI conseguiu se registrar porque o carregador de registros estava removendo a `user` pasta vazia abaixo. `/var/log/hive`
- Corrigido um problema com as importações do Hive, que afetava o particionamento e resultava em um erro durante a importação.
- Corrigido um problema em que EMRFS o `s3-dist-cp` não manipulava adequadamente os nomes de buckets que continham pontos.
- Foi alterado um comportamento EMRFS para que, em buckets com versionamento ativado, o arquivo `_${folder}` marcador não seja criado continuamente, o que pode contribuir para melhorar o desempenho de buckets com versionamento ativado.
- Alterou o comportamento de EMRFS forma que não use arquivos de instruções, exceto nos casos em que a criptografia do lado do cliente está habilitada. Se desejar excluir os arquivos de instrução enquanto usa a criptografia no lado do cliente, você poderá definir a propriedade `emrfs-site.xml fs.s3.cse.cryptoStorageMode.deleteInstructionFiles.enabled`, como verdadeiro.
- A agregação de YARN registros foi alterada para reter os registros no destino da agregação por dois dias. O destino padrão é o HDFS armazenamento do seu cluster. Se desejar mudar essa duração, altere o valor de `yarn.log-aggregation.retain-seconds` usando a classificação

de configuração `yarn-site` quando criar seu cluster. Como sempre, você pode salvar os logs de aplicações no Amazon S3 usando o parâmetro `log-uri` quando criar o cluster.

Patches aplicados

Os seguintes patches de projetos de código aberto foram incluídos nesta versão:

- [HIVE-965](#)
- [HIVE-9183](#)
- [HADOOP-12810](#)

Versão 4.3.0

Data do release: 19 de janeiro de 2016

Recursos

Os seguintes recursos estão disponíveis nesta versão:

- Atualizado para Hadoop 2.7.1
- Atualizado para Spark 1.6.0
- Ganglia atualizado para 3.7.2
- Presto atualizado para 0.130

A Amazon EMR fez algumas alterações em `spark.dynamicAllocation.enabled` quando está definido como verdadeiro; por padrão, é falso. Quando definida como verdadeiro, isso afeta os padrões definidos pela configuração `maximizeResourceAllocation`:

- Se `spark.dynamicAllocation.enabled` estiver definida como `true`, `spark.executor.instances` não será definida por `maximizeResourceAllocation`.
- A configuração `spark.driver.memory` agora é definida com base nos tipos de instância no cluster de maneira semelhante a como `spark.executors.memory` é definida. No entanto, como o aplicativo do driver Spark pode ser executado na instância principal ou em uma das instâncias principais (por exemplo, nos modos YARN cliente e cluster, respectivamente), a `spark.driver.memory` configuração é definida com base no tipo de instância do tipo de instância menor entre esses dois grupos de instâncias.

- A `spark.default.parallelism` configuração agora está definida como o dobro do número de CPU núcleos disponíveis para YARN contêineres. Em versões anteriores, era a metade desse valor.
- Os cálculos da sobrecarga de memória reservada para YARN os processos do Spark foram ajustados para serem mais precisos, resultando em um pequeno aumento na quantidade total de memória disponível para o Spark (ou seja,). `spark.executor.memory`

Problemas conhecidos das versões anteriores que foram resolvidos

- YARNa agregação de registros agora está ativada por padrão.
- Corrigido um problema em que os registros não eram enviados para o bucket de registros do Amazon S3 de um cluster quando a agregação de registros YARN estava ativada.
- YARNos tamanhos de contêineres agora têm um novo mínimo de 32 em todos os tipos de nós.
- Corrigido um problema com o Ganglia que causava E/S de disco excessivas no nó principal em clusters grandes.
- Corrigido um problema que impedia que os logs de aplicações fossem enviados para o Amazon S3 quando um cluster estivesse sendo encerrado.
- Corrigido um problema EMRFS CLI que fazia com que certos comandos falhassem.
- Corrigido um problema com o Zeppelin que impedia que dependências fossem carregadas no subjacente. `SparkContext`
- Corrigido um problema resultante da emissão de um redimensionamento para tentar adicionar instâncias.
- Corrigido um problema no Hive em que o `CREATE TABLE AS SELECT` fazia chamadas de lista excessivas para o Amazon S3.
- Corrigido um problema em que clusters grandes não provisionavam corretamente quando o Hue, o Oozie e o Ganglia estivessem instalados.
- Corrigido um problema no `s3-dist-cp` em que retornava um código de saída zero, mesmo se houvesse falha com um erro.

Patches aplicados

Os seguintes patches de projetos de código aberto foram incluídos nesta versão:

- [OOZIE-2402](#)

- [HIVE-12502](#)
- [HIVE-10 631](#)
- [HIVE-1213](#)
- [HIVE-10 559](#)
- [HIVE-12715](#)
- [HIVE-10 685](#)

Versão 4.2.0

Data do release: 18 de novembro de 2015

Recursos

Os seguintes recursos estão disponíveis nesta versão:

- Adicionado o suporte ao Ganglia
- Atualizado para Spark 1.5.2
- Atualizado para Presto 0.125
- Oozie atualizado para 4.2.0
- Zeppelin atualizado para 0.5.5
- Atualizou o AWS SDK for Java para 1.10.27

Problemas conhecidos das versões anteriores que foram resolvidos

- Corrigido um problema EMRFS CLI em que ele não usava o nome padrão da tabela de metadados.
- Corrigido um problema encontrado ao usar tabelas ORC apoiadas por -back no Amazon S3.
- Corrigido um problema encontrado com uma divergência de versão do Python na configuração do Spark.
- Corrigido um problema quando o status de um YARN nó não era reportado devido a DNS problemas com clusters em umVPC.
- Corrigido um problema encontrado ao YARN descomissionar nós, resultando em aplicativos interrompidos ou na incapacidade de programar novos aplicativos.

- Corrigido um problema encontrado quando os clusters terminavam com o status TIMED _ OUT _STARTING.
- Corrigido um problema encontrado ao incluir a dependência do EMRFS Scala em outras compilações. A dependência Scala foi removida.

Configurar aplicações

Para substituir as configurações padrão de uma aplicação, você pode fornecer um objeto de configuração. Você pode usar uma sintaxe abreviada para fornecer a configuração ou fazer referência ao objeto de configuração em um arquivo. JSON Os objetos de configuração consistem em uma classificação, propriedades e configurações opcionais aninhadas As propriedades correspondem às configurações da aplicação que você deseja alterar. Você pode especificar várias classificações para vários aplicativos em um único JSON objeto.

Warning

APIs operações Amazon EMR Describe e List emitem configurações personalizadas e configuráveis, que são usadas como parte dos fluxos de EMR trabalho da Amazon, em texto simples. Para fornecer informações sigilosas, como senhas, nessas configurações, consulte [Store sensitive configuration data in AWS Secrets Manager](#).

As classificações de configuração disponíveis variam de acordo com a versão de EMR lançamento da Amazon. Para obter uma lista das classificações de configuração compatíveis com determinada versão, consulte a página da versão em [Sobre os EMR lançamentos da Amazon](#).

Veja a seguir um exemplo de JSON arquivo para uma lista de configurações.

```
[
  {
    "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"
    }
  }
]
```


Uma classificação de configuração geralmente é mapeada para um arquivo de configuração específico do aplicativo. Por exemplo, a classificação `hive-site` é mapeada para as configurações no arquivo de configuração `hive-site.xml` para o Hive. Uma exceção para isso é a ação de bootstrap `configure-daemons`, que já não tem mais suporte, que é usada para definir parâmetros de ambiente, como `--namenode-heap-size`. Opções como esta são incluídas nas classificações `hadoop-env` e `yarn-env` com suas próprias classificações de exportação aninhadas. Se qualquer classificação terminar com `env`, use a subclassificação de exportação.

Outra exceção é `s3get`, que é usado para colocar um objeto `EncryptionMaterialsProvider` cliente em cada nó de um cluster para uso na criptografia do lado do cliente. Uma opção foi adicionada à classificação `emrfs-site` para essa finalidade.

Veja a seguir um exemplo da classificação `hadoop-env`.

```
[
  {
    "Classification": "hadoop-env",
    "Properties": {

    },
    "Configurations": [
      {
        "Classification": "export",
        "Properties": {
          "HADOOP_DATANODE_HEAPSIZE": "2048",
          "HADOOP_NAMENODE_OPTS": "-XX:GCTimeRatio=19"
        },
        "Configurations": [

        ]
      }
    ]
  }
]
```

Veja a seguir um exemplo da classificação `yarn-env`.

```
[
  {
    "Classification": "yarn-env",
    "Properties": {
```

```

    },
    "Configurations": [
      {
        "Classification": "export",
        "Properties": {
          "YARN_RESOURCEMANAGER_OPTS": "-Xdebug -Xrunjdp:transport=dt_socket"
        },
        "Configurations": [
          ]
        ]
      }
    ]
  }
]

```

As configurações a seguir não pertencem a um arquivo de configuração, mas são usadas pela Amazon EMR para potencialmente definir várias configurações em seu nome.

Configurações selecionadas pela Amazon EMR

Aplicativo	Classificação do rótulo da versão	Propriedades válidas	Quando usar
Spark	spark	maximizeResourceAllocation	Configure executores para utilizar o máximo de recursos de cada nó.

Tópicos

- [Configurar aplicações ao criar um cluster](#)
- [Reconfigurar um grupo de instâncias de um cluster em execução](#)
- [Armazenar dados de configuração sigilosos no AWS Secrets Manager](#)
- [Configurar aplicações para usar uma Máquina Virtual Java específica](#)

Configurar aplicações ao criar um cluster

Ao criar um cluster, você pode substituir as configurações padrão dos aplicativos usando o EMR console da Amazon, o AWS Command Line Interface (AWS CLI) ou o AWS SDK

Para substituir a configuração padrão de uma aplicação, especifique valores personalizados em uma classificação de configuração. Uma classificação de configuração corresponde a um XML arquivo de configuração de um aplicativo, como `hive-site.xml`.

As classificações de configuração variam de acordo com a versão de EMR lançamento da Amazon. Para obter uma lista de classificações de configuração que estão disponíveis em uma versão específica, consulte a página de detalhes da versão. Por exemplo, a [EMR versão 6.4.0 da Amazon](#).

Fornecer uma configuração no console ao criar um cluster

Para fornecer uma configuração, navegue até a página Criar cluster e expanda Configurações de software. Em seguida, você pode inserir a configuração diretamente usando uma JSON ou uma sintaxe abreviada demonstrada em texto sombreado no console. Caso contrário, você pode fornecer um Amazon S3 URI para um arquivo com um JSON Configurations objeto.

Para fornecer uma configuração para um grupo de instâncias, escolha um cluster na sua lista de clusters e, em seguida, escolha a guia Configurações. Na tabela de configurações do grupo de instâncias, escolha o grupo de instâncias a ser editado e, em seguida, escolha Reconfigurar.

Forneça uma configuração usando o AWS CLI ao criar um cluster

Você pode fornecer uma configuração `create-cluster` fornecendo um caminho para um JSON arquivo armazenado localmente ou no Amazon S3. O exemplo a seguir pressupõe que você esteja usando funções padrão para a Amazon EMR e que as funções tenham sido criadas. Se você precisar criar as funções, execute `aws emr create-default-roles` primeiro.

Se a configuração estiver em seu diretório local, você poderá usar o exemplo de comando a seguir.

```
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
```

Se a configuração estiver em um caminho do Amazon S3, você precisará configurar a seguinte solução alternativa antes de passar o caminho do Amazon S3 para o comando `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
```

Forneça uma configuração usando o Java SDK ao criar um cluster

O trecho de programa a seguir mostra como fornecer uma configuração usando o 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")  
);
```

Reconfigurar um grupo de instâncias de um cluster em execução

Com a Amazon EMR versão 5.21.0 e posterior, você pode reconfigurar aplicativos de cluster e especificar classificações de configuração adicionais para cada grupo de instâncias em um cluster

em execução. Para fazer isso, você pode usar o EMR console da Amazon, o AWS Command Line Interface (AWS CLI) ou AWS SDK o.

Quando você atualiza uma configuração de aplicativo para um grupo de instâncias no novo EMR console da Amazon, o console tenta mesclar a nova configuração com a configuração existente para criar uma configuração nova e ativa. No caso incomum em que a Amazon não EMR consegue mesclar a configuração, o console alerta você.

Depois de enviar uma solicitação de reconfiguração para um grupo de instâncias, a Amazon EMR atribui um número de versão à nova especificação de configuração. Você pode acompanhar o número da versão de uma configuração ou o estado de um grupo de instâncias visualizando os CloudWatch eventos. Para obter mais informações, consulte [Monitorar CloudWatch eventos](#).

Note

Só é possível substituir, e não excluir, as configurações do cluster que foram especificadas durante a criação do cluster. Se houver diferenças entre a configuração existente e o arquivo que você fornece, a Amazon EMR redefine as configurações modificadas manualmente, como as que você modificou enquanto estava conectado ao seu cluster usando SSH, para os padrões do cluster para o grupo de instâncias especificado.

Considerações ao reconfigurar um grupo de instâncias

Ações de reconfiguração


Quando você envia uma solicitação de reconfiguração usando o EMR console da Amazon, o AWS Command Line Interface (AWS CLI) ou o AWS SDK, a Amazon EMR verifica o arquivo de configuração existente no cluster. Se houver diferenças entre a configuração existente e o arquivo que você fornece, a Amazon EMR inicia ações de reconfiguração, reinicia alguns aplicativos e redefine todas as configurações modificadas manualmente, como as que você modificou enquanto estava conectado ao seu cluster usando SSH, de acordo com os padrões do cluster para o grupo de instâncias especificado.

Note

A Amazon EMR executa algumas ações padrão durante cada reconfiguração do grupo de instâncias. Essas ações padrão podem entrar em conflito com as personalizações

do cluster feitas por você e resultar em falhas de reconfiguração. Para obter mais informações sobre como solucionar falhas de reconfiguração, consulte [Solucionar problemas de reconfiguração do grupo de instâncias](#).

A Amazon EMR também inicia ações de reconfiguração para as classificações de configuração que você especifica em sua solicitação. Para obter uma lista completa dessas ações, consulte a seção [Classificações de configuração da versão da Amazon EMR que você usa](#). Por exemplo, [Classificações de configuração do 6.2.0](#).

 Note

O Amazon EMR Release Guide lista apenas as ações de reconfiguração a partir das EMR versões 5.32.0 e 6.2.0 da Amazon.

Interrupção do serviço

A Amazon EMR segue um processo contínuo para reconfigurar instâncias nos grupos de instâncias Task e Core. Somente 10% das instâncias em um grupo de instâncias são modificadas e reiniciadas por vez. Esse processo leva mais tempo para ser concluído, mas reduz a chance de possíveis falhas de aplicativo em um cluster em execução.

Para executar YARN trabalhos durante uma YARN reinicialização, você pode criar um EMR cluster da Amazon com vários nós principais ou `yarn.resourcemanager.recovery.enabled` definir isso `true` na sua classificação de `yarn-site` configuração. Para obter mais informações sobre o uso de vários nós principais, consulte [Alta disponibilidade YARN ResourceManager](#).

Validação da aplicação

A Amazon EMR verifica se cada aplicativo no cluster está em execução após o processo de reinicialização da reconfiguração. Se alguma aplicação não estiver disponível, a operação geral de reconfiguração falhará. Se uma operação de reconfiguração falhar, a Amazon EMR reverte os parâmetros de configuração para a versão de trabalho anterior.

Note

Para evitar falhas na reconfiguração, é recomendável instalar somente as aplicações que planeja usar no cluster. Também é recomendável verificar se todas as aplicações do cluster estão íntegras e em execução antes de enviar uma solicitação de reconfiguração.

Tipos de reconfiguração

É possível reconfigurar um grupo de instâncias de duas maneiras:

- **Substituir.** Método de reconfiguração padrão e o único disponível nas EMR versões anteriores à 5.35.0 e 6.6.0 da Amazon. Esse método de reconfiguração substitui indiscriminadamente todo arquivo no cluster pelo conjunto de configurações recém-enviado. O método apaga todas as alterações nos arquivos de configuração feitas fora da API reconfiguração.
- **Mesclar.** Método de reconfiguração compatível com as EMR versões 5.35.0 e 6.6.0 e posteriores da Amazon, exceto no EMR console da Amazon, onde nenhuma versão é compatível. Esse método de reconfiguração mescla as configurações recém-enviadas com as configurações já existentes no cluster. Essa opção só adicionará ou modificará as novas configurações que você enviar. Ela preserva as configurações atuais.

Note

EMRA Amazon continua substituindo algumas configurações essenciais do Hadoop necessárias para garantir que o serviço esteja funcionando corretamente.

Limitações

Ao reconfigurar um grupo de instâncias de um cluster em execução, considere estas limitações:

- YARNs não-aplicativos podem falhar durante a reinicialização ou causar problemas no cluster, especialmente se os aplicativos não estiverem configurados corretamente. Clusters que se aproximam do máximo de memória e CPU uso podem ter problemas após o processo de reinicialização. Isso se aplica principalmente ao grupo de instâncias principal.
- Não é possível enviar uma solicitação de reconfiguração quando o grupo de instâncias está sendo redimensionado. Se uma reconfiguração for iniciada enquanto um grupo de instâncias estiver sendo redimensionado, a reconfiguração não poderá ser iniciada até que o redimensionamento do grupo de instâncias tenha sido concluído e vice-versa.

- Depois de reconfigurar um grupo de instâncias, a Amazon EMR reinicia os aplicativos para permitir que as novas configurações entrem em vigor. Uma falha no trabalho ou outro comportamento inesperado do aplicativo poderão ocorrer se os aplicativos estiverem em uso durante a reconfiguração.
- Se a reconfiguração de um grupo de instâncias falhar, a Amazon EMR reverte os parâmetros de configuração para a versão de trabalho anterior. Se também houver falha no processo de reversão, você deverá enviar uma nova solicitação `ModifyInstanceGroup` para recuperar o grupo de instâncias a partir do estado `SUSPENDED`.
- Solicitações de reconfiguração para classificações de configuração do Phoenix são suportadas somente na Amazon EMR versão 5.23.0 e posterior, e não são suportadas na Amazon EMR versão 5.21.0 ou 5.22.0.
- Solicitações de HBase reconfiguração para classificações de configuração são suportadas somente na Amazon EMR versão 5.30.0 e posterior, e não são suportadas nas EMR versões 5.23.0 a 5.29.0 da Amazon.
- A Amazon EMR oferece suporte a solicitações de reconfiguração de aplicativos em um EMR cluster da Amazon com vários nós primários somente nas EMR versões 5.27.0 e posteriores da Amazon.
- A reconfiguração da `hdfs-encryption-zones` classificação ou de qualquer uma das classificações de KMS configuração do Hadoop não é suportada em um EMR cluster da Amazon com vários nós primários.
- EMR Atualmente, a Amazon não oferece suporte a determinadas solicitações de reconfiguração para o agendador de capacidade que exijam a reinicialização do YARN ResourceManager Por exemplo, não é possível remover uma fila completamente.

Reconfigurar um grupo de instâncias no console

Note

O EMR console da Amazon não suporta reconfigurações do tipo Merge.

1. Abra o EMR console da Amazon em <https://console.aws.amazon.com/emr>
2. Na lista de clusters, m Nome, escolha o cluster ativo que você deseja reconfigurar.
3. Abra a página de detalhes do cluster e acesse a guia Configurações.
4. Na lista suspensa Filter (Filtro), selecione o grupo de instâncias que você deseja reconfigurar.

5. No menu suspenso Reconfigurar, escolha Editar na tabela ou Editar no arquivo. JSON
 - Editar na tabela: na tabela de classificações de configuração, edite a propriedade e o valor das configurações atuais ou escolha Adicionar configuração para fornecer a propriedade e o valor para outras classificações de configuração.
 - Editar no JSON arquivo - insira a configuração diretamente ou use a sintaxe abreviada (demonstrada em texto sombreado). JSON Caso contrário, forneça um Amazon S3 URI para um arquivo com um JSON Configurations objeto.

Note

A coluna Origem na tabela de classificações de configuração indica se a configuração é fornecida quando você cria um cluster ou quando você especifica configurações adicionais para esse grupo de instâncias. Você pode editar as configurações para um grupo de instâncias de ambas as origens. Não é possível excluir as configurações iniciais do cluster, mas você pode substituí-las por um grupo de instâncias. Você também pode adicionar ou editar classificações de configuração aninhadas diretamente na tabela. Por exemplo, para fornecer uma sub-classificação de `export` adicional de `hadoop-env`, adicione uma classificação de configuração `hadoop.export` na tabela. Em seguida, forneça uma propriedade específica e um valor para essa classificação.

6. (Opcional) Selecione Apply this configuration to all active instance groups (Aplicar essa configuração a todos os grupos de instâncias ativas).
7. Salve as alterações.

Reconfigure um grupo de instâncias usando o CLI

Você pode usar o comando `modify-instance-groups` para especificar configurações para cada grupo de instâncias em um cluster em execução.

Note

Nos exemplos a seguir, substitua `<j-2AL4XXXXXX5T9>` com seu ID de cluster e substitua `<ig-1xxxxxxx9>` com seu ID de grupo de instâncias.

Example : substituir uma configuração para um grupo de instâncias

O exemplo a seguir faz referência a um JSON arquivo de configuração chamado `instanceGroups.json` para editar a propriedade do verificador de integridade do YARN NodeManager disco para um grupo de instâncias.

1. Prepare a classificação da configuração e salve-a como `instanceGroups.json` no mesmo diretório em que você executará o comando.

```
[
  {
    "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. Execute o seguinte comando .

```
aws emr modify-instance-groups --cluster-id <j-2AL4XXXXXX5T9> \
--instance-groups file://instanceGroups.json
```

Example — Adicionar uma configuração a um grupo de instâncias

Se deseja adicionar uma configuração a um grupo de instâncias, você deverá incluir todas as configurações especificadas anteriormente para aquele grupo de instâncias em sua nova solicitação `ModifyInstanceGroup`. Caso contrário, as configurações especificadas anteriormente serão removidas.

O exemplo a seguir adiciona uma propriedade para o verificador de memória YARN NodeManager virtual. A configuração também inclui valores previamente especificados para o verificador de integridade do YARN NodeManager disco, para que os valores não sejam substituídos.

1. Prepare o conteúdo a seguir em `instanceGroups.json` e salve-o no mesmo diretório onde você executará o comando.

```
[
  {
    "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. Execute o seguinte comando .

```
aws emr modify-instance-groups --cluster-id <j-2AL4XXXXXX5T9> \
--instance-groups file://instanceGroups.json
```

Example — Adicionar uma configuração a um grupo de instâncias com uma reconfiguração do tipo Merge

Quando quiser usar o método padrão de reconfiguração `Overwrite` para adicionar uma configuração, é necessário incluir todas as configurações especificadas anteriormente para esse grupo de instâncias em sua nova solicitação `ModifyInstanceGroup`. Caso contrário, `Overwrite` removerá as configurações que você especificou anteriormente. Não é necessário fazer isso

com a reconfiguração Merge. Em vez disso, garanta que sua solicitação inclua apenas as novas configurações.

O exemplo a seguir adiciona uma propriedade para o verificador de memória YARN NodeManager virtual. Como essa é uma reconfiguração do tipo Merge, ela não sobrescreve os valores especificados anteriormente para o verificador de integridade do YARN NodeManager disco.

1. Prepare o conteúdo a seguir em `instanceGroups.json` e salve-o no mesmo diretório onde você executará o comando.

```
[
  {
    "InstanceGroupId": "<ig-1xxxxxxx9>",
    "ReconfigurationType": "MERGE",
    "Configurations": [
      {
        "Classification": "yarn-site",
        "Properties": {
          "yarn.nodemanager.vmem-check-enabled": "true",
          "yarn.nodemanager.vmem-pmem-ratio": "3.0"
        }
      },
      {
        "Classification": "yarn-site",
        "Properties": {
          "yarn.nodemanager.vmem-check-enabled": "true",
          "yarn.nodemanager.vmem-pmem-ratio": "3.0"
        }
      }
    ]
  }
]
```

2. Execute o seguinte comando .

```
aws emr modify-instance-groups --cluster-id <j-2AL4XXXXXX5T9> \
--instance-groups file://instanceGroups.json
```

Example : excluir uma configuração de um grupo de instâncias

Para excluir uma configuração para um grupo de instâncias, envie uma nova solicitação de reconfiguração que exclua a configuração anterior.

Note

Só é possível substituir a configuração inicial do cluster. Você não pode excluí-la.

Por exemplo, para excluir a configuração do verificador de integridade do YARN NodeManager disco do exemplo anterior, envie uma nova `instanceGroups.json` com o conteúdo a seguir.

```
[
  {
    "InstanceId": "<ig-1xxxxxxx9>",
    "Configurations": [
      {
        "Classification": "yarn-site",
        "Properties": {
          "yarn.nodemanager.vmem-check-enabled": "true",
          "yarn.nodemanager.vmem-pmem-ratio": "3.0"
        },
        "Configurations": []
      }
    ]
  }
]
```

Note

Para excluir todas as configurações na última solicitação de reconfiguração, envie uma solicitação de reconfiguração com uma matriz de configurações. Por exemplo,

```
[
  {
    "InstanceId": "<ig-1xxxxxxx9>",
    "Configurations": []
  }
]
```

Example — Reconfigurar e redimensionar um grupo de instâncias em uma solicitação

O exemplo a seguir JSON demonstra como reconfigurar e redimensionar um grupo de instâncias na mesma solicitação.

```
[
```

```

{
  "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"
      }
    }
  ]
}
]

```

Reconfigurar um grupo de instâncias usando o Java SDK

Note

Nos exemplos a seguir, substitua `<j-2AL4XXXXXX5T9>` com seu ID de cluster e substitua `<ig-1xxxxxxx9>` com seu ID de grupo de instâncias.

O trecho de código a seguir fornece uma nova configuração para um grupo de instâncias usando o 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);

```

O trecho de código a seguir exclui uma configuração especificada anteriormente para um grupo de instâncias fornecendo uma matriz vazia de configurações.

```

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);

```

Solucionar problemas de reconfiguração do grupo de instâncias

Se o processo de reconfiguração de um grupo de instâncias falhar, a Amazon EMR reverterá a reconfiguração e registrará uma mensagem de falha usando um evento da Amazon CloudWatch. O evento fornece um breve resumo da falha de reconfiguração. Ele lista as instâncias em que a

reconfiguração falhou e as mensagens de falha correspondentes. Veja a seguir um exemplo de mensagem de falha.

```
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".
```

Para coletar mais dados sobre uma falha de reconfiguração, verifique os logs de provisionamento do nó. Isso serve principalmente quando você recebe a mensagem a seguir.

```
i-xxxxxxx1 failed with message "Unable to complete transaction and some changes were applied."
```

On the node

Acessar os logs de provisionamento do nó conectando-se a um nó

1. Use SSH para se conectar ao nó no qual a reconfiguração falhou. Para obter instruções, consulte [Connect to your Linux instance](#) no Amazon EC2 User Guide for Linux Instances.
2. Navegue até o diretório a seguir, que contém os arquivos de log de provisionamento do nó.

```
/mnt/var/log/provision-node/
```

3. Abra o subdiretório `reports` e pesquise o relatório de provisionamento do nó para a reconfiguração. O `reports` diretório organiza os registros por número da versão de reconfiguração, identificador universal exclusivo (UUID), endereço IP da EC2 instância Amazon e data e hora. Cada relatório é um YAML arquivo compactado que contém informações detalhadas sobre o processo de reconfiguração.

Veja a seguir um exemplo de nome e caminho do arquivo de relatório.

```
/reports/2/ca598xxx-cxxx-4xxx-bxxx-6dbxxxxxxxxx/ip-10-73-xxx-xxx.ec2.internal/202104061715.yaml.gz
```

4. Você pode examinar um relatório usando um visualizador de arquivos como `zless`, como no exemplo a seguir.


```
zless 202104061715.yaml.gz
```

Amazon S3

Acessar logs de provisionamento do nó usando o Amazon S3

1. Faça login no AWS Management Console e abra o console do Amazon S3 em. <https://console.aws.amazon.com/s3/>
2. Abra o bucket do Amazon S3 especificado quando você configurou o cluster para arquivar arquivos de log.
3. Navegue até esta pasta, que contém os arquivos de log de provisionamento do nó:

```
DOC-EXAMPLE-BUCKET/elasticmapreduce/<cluster id>/node/<instance id>/provision-  
node/
```

4. Abra a pasta `reports` e pesquise o relatório de provisionamento do nó para a reconfiguração. A `reports` pasta organiza os registros por número da versão de reconfiguração, identificador universal exclusivo (UUID), endereço IP da EC2 instância Amazon e data e hora. Cada relatório é um YAML arquivo compactado que contém informações detalhadas sobre o processo de reconfiguração.

Veja a seguir um exemplo de nome e caminho do arquivo de relatório.

```
/reports/2/ca598xxx-cxxx-4xxx-bxxx-6dbxxxxxxxxxx/ip-10-73-xxx-  
xxx.ec2.internal/202104061715.yaml.gz
```

5. Para visualizar um arquivo de log, você pode baixá-lo no Amazon S3 para a máquina local como arquivo de texto. Para obter instruções, consulte [Fazer download de um objeto](#).

Cada arquivo de log contém um relatório detalhado de provisionamento para a reconfiguração associada. Para encontrar informações sobre mensagens de erro, pesquise o nível do log `err` de um relatório. O formato do relatório depende da versão da Amazon EMR em seu cluster.

O exemplo a seguir mostra informações de erro para versões de EMR lançamento da Amazon anteriores à 5.32.0 e 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
```

Em vez disso, as EMR versões 5.32.0 e 6.2.0 e posteriores da Amazon usam o seguinte formato.

```
- level: err
  message: 'Example detailed error message.'
  source: Puppet
  tags:
  - err
  time: '2021-01-01 00:00:00.000000 +00:00'
  file:
  line:
```

Armazenar dados de configuração sigilosos no AWS Secrets Manager

A Amazon EMR descreve e lista API as operações que emitem dados de configuração personalizados (como `DescribeCluster` e `ListInstanceGroups`) em texto simples. A Amazon EMR se integra AWS Secrets Manager para que você possa armazenar seus dados no Secrets Manager e usar o segredo ARN em suas configurações. Dessa forma, você não passa dados de configuração confidenciais para a Amazon EMR em texto sem formatação e os expõe ao ambiente externo. APIs Se você indicar que um par de valores-chave contém um ARN para um segredo armazenado no Secrets Manager, a Amazon EMR recupera esse segredo ao enviar dados de configuração para o cluster. A Amazon EMR não envia a anotação quando usa o externo APIs para exibir a configuração.

Criar um segredo

Para criar um segredo, siga as instruções em [Criação de um segredo do AWS Secrets Manager](#) no Guia do usuário do AWS Secrets Manager . Na Etapa 3, você deve escolher o campo Texto sem formatação para inserir o valor sigiloso.

Observe que, embora o Secrets Manager permita que um segredo contenha até 65536 bytes, a Amazon EMR limita o tamanho combinado da chave de propriedade (excluindo a anotação) e o valor secreto recuperado a 1024 caracteres.

Conceda EMR à Amazon acesso para recuperar o segredo

A Amazon EMR usa uma função IAM de serviço para provisionar e gerenciar clusters para você. A função de serviço da Amazon EMR define as ações permitidas para a Amazon EMR quando ela provisiona recursos e executa tarefas de nível de serviço que não são executadas no contexto de uma EC2 instância da Amazon em execução em um cluster. Para obter mais informações sobre funções de serviço, consulte [Função de serviço para Amazon EMR \(EMRfunção\)](#) e [Personalizar IAM funções](#).

Para permitir que EMR a Amazon recupere o valor secreto do Secrets Manager, adicione a seguinte declaração de política à sua EMR função na Amazon ao iniciar seu cluster.

```
{
  "Sid": "AllowSecretsRetrieval",
  "Effect": "Allow",
  "Action": "secretsmanager:GetSecretValue",
  "Resource": [
    "arn:aws:secretsmanager:<region>:<aws-account-id>:secret:<secret-name>"
  ]
}
```

Se você criar o segredo com uma chave gerenciada pelo cliente AWS KMS key, também deverá adicionar `kms:Decrypt` permissão à EMR função da Amazon para a chave que você usa. Para obter mais informações, consulte [Controle de acesso e autenticação para o AWS Secrets Manager](#) no Manual do usuário do AWS Secrets Manager .

Usar o segredo em uma classificação de configuração

Você pode adicionar a `EMR.secret@` anotação a qualquer propriedade de configuração para indicar que seu par de valores-chave contém um ARN para um segredo armazenado no Secrets Manager.

O exemplo a seguir mostra como fornecer um segredo ARN em uma classificação de configuração:

```
{
  "Classification": "core-site",
  "Properties": {
```

```

    "prestos3.access-key": "<sensitive-access-key>",
    "EMR.secret@prestos3.secret-key": "arn:aws:secretsmanager:<region>:<aws-account-id>:secret:<secret-name>"
  }
}

```

Quando você cria seu cluster e envia sua configuração anotada, a Amazon EMR valida as propriedades de configuração. Se sua configuração for válida, a Amazon EMR retira a anotação da configuração e recupera o segredo do Secrets Manager para criar a configuração real antes de aplicá-la ao cluster:

```

{
  "Classification": "core-site",
  "Properties": {
    "prestos3.access-key": "<sensitive-access-key>",
    "prestos3.secret-key": "<my-secret-key-retrieved-from-Secrets-Manager>"
  }
}

```

Quando você chama uma ação como `DescribeCluster`, a Amazon EMR retorna a configuração atual do aplicativo no cluster. Se uma propriedade de configuração do aplicativo estiver marcada como contendo um segredo ARN, a configuração do aplicativo retornada pela `DescribeCluster` chamada conterá o valor ARN e não o segredo. Isso garante que o valor do segredo seja visível somente no cluster:

```

{
  "Classification": "core-site",
  "Properties": {
    "prestos3.access-key": "<sensitive-access-key>",
    "prestos3.secret-key": "arn:aws:secretsmanager:<region>:<aws-account-id>:secret:<secret-name>"
  }
}

```

Atualizar o valor do segredo

A Amazon EMR recupera o valor secreto de uma configuração anotada sempre que o grupo de instâncias anexado estiver iniciando, reconfigurando ou redimensionando. É possível usar o Secrets Manager para modificar o valor de um segredo utilizado na configuração de um cluster em execução. Ao fazer isso, você pode enviar uma solicitação de reconfiguração para cada grupo de instâncias

que deve receber o valor atualizado. Para obter mais informações sobre como reconfigurar um grupo de instâncias e o que considerar ao fazer isso, consulte [Reconfigurar um grupo de instâncias de um cluster em execução](#).

Configurar aplicações para usar uma Máquina Virtual Java específica

As EMR versões da Amazon têm diferentes versões padrão do Java Virtual Machine (JVM). Esta página explica o JVM suporte para diferentes versões e aplicativos.

Considerações

Para obter informações sobre as versões Java suportadas para aplicativos, consulte as páginas do aplicativo no [Amazon EMR Release Guide](#).

- A Amazon EMR só suporta a execução de uma versão de tempo de execução em um cluster e não suporta a execução de diferentes nós ou aplicativos em diferentes versões de tempo de execução no mesmo cluster.
- Para o Amazon EMR 7.x, a Java Virtual Machine (JVM) padrão é o Java 17 para aplicativos que oferecem suporte ao Java 17, com exceção do Apache Livy. Para obter mais informações sobre as JDK versões suportadas dos aplicativos, consulte a página de lançamento correspondente no Amazon EMR Release Guide.
- A partir do Amazon EMR 7.1.0, o Flink suporta e está configurado para Java 17 por padrão. Para usar uma versão diferente do Java Runtime, substitua as configurações em `flink-conf`. Para obter mais informações sobre como configurar o Flink para usar o Java 8 ou o Java 11, consulte [Configurar o Flink para ser executado com](#) o Java 11.
- Para Amazon EMR 5.x e 6.x, a Java Virtual Machine (JVM) padrão é Java 8.
 - Para as EMR versões 6.12.0 e superiores da Amazon, alguns aplicativos também oferecem suporte ao Java 11 e 17.
 - Para as EMR versões 6.9.0 e superiores da Amazon, o Trino oferece suporte ao Java 17 como padrão. Para obter mais informações sobre o Java 17 com Trino, consulte [Trino updates to Java 17](#) no blog do Trino.

Lembre-se das seguintes considerações específicas da aplicação ao escolher sua versão de runtime:

Notas de configuração Java específicas da aplicação

Aplicativo	Notas sobre configurações do Java
Spark	<p>Para executar o Spark com uma versão Java que não seja padrão, é necessário configurar o Spark e o Hadoop. Para ver exemplos, consulte Substitua o JVM.</p> <ul style="list-style-type: none"> Configure <code>JAVA_HOME</code> em <code>spark-env</code> para atualizar o runtime Java dos processos da instância primária. Por exemplo, <code>spark-submit</code>, <code>spark-shell</code> e Spark History Server. Modifique a configuração do Hadoop para atualizar o tempo de execução Java dos executores do Spark e do YARN ApplicationMaster
Faísca RAPIDS	Você pode executar RAPIDS com a versão Java configurada para o Spark.
Iceberg	Você pode executar o Iceberg com a versão Java configurada da aplicação que o está usando.
Delta	Você pode executar o Delta com a versão Java configurada da aplicação que o está usando.
Hudi	Você pode executar o Hudi com a versão Java configurada da aplicação que o está usando.
Hadoop	Para atualizar o JVM para o Hadoop, modifique <code>.hadoop-env</code> Para ver exemplos, consulte Substitua o JVM .

Aplicativo	Notas sobre configurações do Java
Hive	Para definir a versão Java como 11 ou 17 para o Hive, defina a JVM configuração do Hadoop para a versão Java que você deseja usar.
HBase	Para atualizar o JVM formulário HBase, modifique <code>hbase-env</code> . Por padrão, a Amazon EMR define o HBase JVM com base na JVM configuração do Hadoop, a menos que você substitua as configurações em <code>hbase-env</code> . Para ver exemplos, consulte Substitua o JVM .
Flink	Para atualizar o JVM para o Flink, modifique <code>flink-conf</code> . Por padrão, a Amazon EMR define o Flink JVM com base na JVM configuração do Hadoop, a menos que você substitua as configurações em <code>flink-conf</code> . Para ter mais informações, consulte Configurar o Flink para ser executado com o Java 11 .
Oozie	Para configurar o Oozie para ser executado no Java 11 ou 17, configure o Oozie Server, o Oozie Launcher AM Launcher AM e altere as configurações de trabalho e executáveis do lado do cliente. Também é possível configurar <code>EmbeddedOozieServer</code> para ser executado no Java 17. Para ter mais informações, consulte Configurar a versão do Java para Oozie .
Pig	O Pig é compatível apenas com Java 8. Não é possível usar o Java 11 ou 17 com o Hadoop e executar o Pig no mesmo cluster.

Substitua o JVM

Para substituir a JVM configuração de uma EMR versão da Amazon — por exemplo, para usar o Java 17 com um cluster que usa a EMR versão 6.12.0 da Amazon — forneça a JAVA_HOME configuração para sua classificação de ambiente, que é *application*-env para todos os aplicativos, exceto o Flink. Para o Flink, a classificação do ambiente é *flink-conf*. Para obter as etapas para configurar o runtime Java com o Flink, consulte [Configurar o Flink para ser executado com o Java 11](#).

Tópicos

- [Substitua a JVM configuração com o Apache Spark](#)
- [Substitua a JVM configuração com o Apache HBase](#)
- [Substitua a JVM configuração com o Apache Hadoop e o Hive](#)

Substitua a JVM configuração com o Apache Spark

Quando você usa o Spark com as EMR versões 6.12 e posteriores da Amazon, se você escrever um driver para envio no modo de cluster, o driver usa o Java 8, mas você pode configurar o ambiente para que os executores usem o Java 11 ou 17. E quando você usa o Spark com EMR versões da Amazon inferiores a 5.x e escreve um driver para envio no modo de cluster, o driver usa o Java 7. No entanto, você pode configurar o ambiente para garantir que os executores usem o Java 8.

Para substituir o do JVM Spark, recomendamos que você defina as classificações do Hadoop e do 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": {}
    }
  ]
}

```

Substitua a JVM configuração com o Apache HBase

Para configurar HBase para usar o Java 11, você pode definir a seguinte configuração ao iniciar o cluster.

```

[
  {
    "Classification": "hbase-env",
    "Configurations": [
      {
        "Classification": "export",
        "Configurations": [],
        "Properties": {
          "JAVA_HOME": "/usr/lib/jvm/jre-11"
        }
      }
    ],
    "Properties": {}
  }
]

```

Substitua a JVM configuração com o Apache Hadoop e o Hive

O exemplo a seguir mostra como definir a versão 17 JVM para Hadoop e Hive.

```

[
  {
    "Classification": "hadoop-env",
    "Configurations": [

```

```

        {
            "Classification": "export",
            "Configurations": [],
            "Properties": {
                "JAVA_HOME": "/usr/lib/jvm/jre-17"
            }
        }
    ],
    "Properties": {}
}
]

```

Portas de serviço

A seguir estão YARN as portas HDFS de serviço. Essas configurações refletem os padrões do Hadoop. Outros serviços de aplicativos são hospedados em portas padrão, a menos que documentado de outra forma. Para obter mais informações, consulte a documentação de projeto do aplicativo.

Configurações de porta para YARN e HDFS

Configuração	Nome do host/Porta
<code>fs.default.name</code>	padrão (<code>hdfs://<i>emrDeterminedIP</i>:8020</code>)
<code>dfs.datanode.address</code>	padrão (<code>0.0.0.0:50010</code>)
<code>dfs.datanode.http.address</code>	padrão (<code>0.0.0.0:50075</code>)
<code>dfs.datanode.https.address</code>	padrão (<code>0.0.0.0:50475</code>)
<code>dfs.datanode.ipc.address</code>	padrão (<code>0.0.0.0:50020</code>)
<code>dfs.http.address</code>	padrão (<code>0.0.0.0:50070</code>)
<code>dfs.https.address</code>	padrão (<code>0.0.0.0:50470</code>)
<code>dfs.secondary.http.address</code>	padrão (<code>0.0.0.0:50090</code>)
<code>yarn.nodemanager.address</code>	padrão (<code>\${yarn.nodemanager.hostname}:0</code>)

Configuração	Nome do host/Porta
<code>yarn.nodemanager.localizer.address</code>	padrão (<code>\${yarn.nodemanager.hostname}:8040</code>)
<code>yarn.nodemanager.webapp.address</code>	padrão (<code>\${yarn.nodemanager.hostname}:8042</code>)
<code>yarn.resourcemanager.address</code>	padrão (<code>\${yarn.resourcemanager.hostname}:8032</code>)
<code>yarn.resourcemanager.admin.address</code>	padrão (<code>\${yarn.resourcemanager.hostname}:8033</code>)
<code>yarn.resourcemanager.resource-tracker.address</code>	padrão (<code>\${yarn.resourcemanager.hostname}:8031</code>)
<code>yarn.resourcemanager.scheduler.address</code>	padrão (<code>\${yarn.resourcemanager.hostname}:8030</code>)
<code>yarn.resourcemanager.webapp.address</code>	padrão (<code>\${yarn.resourcemanager.hostname}:8088</code>)
<code>yarn.web-proxy.address</code>	padrão (no-value)
<code>yarn.resourcemanager.hostname</code>	<i>emrDeterminedIP</i>

Note

O termo *emrDeterminedIP* é um endereço IP gerado pelo plano de EMR controle da Amazon. Na versão mais recente, essa convenção foi eliminada, com exceção para as configurações `yarn.resourcemanager.hostname` e `fs.default.name`.

Usuários do aplicativo

Os aplicativos executarão processos como seus próprios usuários. Por exemplo, o Hive é JVMs executado como usuáriohive, MapReduce JVMs executado como mapred e assim por diante. Isso é demonstrado no exemplo a seguir de status de processo.

```

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

```

Verificar dependências usando o repositório de artefatos do Amazon EMR

Você pode usar o repositório de artefatos do Amazon EMR para compilar o código de trabalho do Apache Hive e do Apache Hadoop relacionado às versões exatas de bibliotecas e dependências disponíveis com versões específicas do Amazon EMR, a partir da versão 5.18.0 do Amazon EMR. A compilação de artefatos do Amazon EMR no repositório ajuda a evitar problemas do caminho de classe do runtime, garantindo que as versões das bibliotecas nas quais o trabalho é criado sejam exatamente as mesmas versões fornecidas no runtime no cluster. Atualmente, os artefatos do Amazon EMR estão disponíveis somente para compilações do Maven.

Para acessar o repositório de artefatos, adicione o URL do repositório ao arquivo de configurações Maven ou ao arquivo de configurações `pom.xml` de um projeto específico. Em seguida, você pode especificar as dependências na configuração de seu projeto. Para as versões de dependência, use a versão listada em Versões do componente para a versão desejada em [Versões de lançamento do Amazon EMR 5.x](#). Por exemplo, as versões de componente para a versão do Amazon EMR mais recente está disponível em [the section called “Versões do componente 5.36.2”](#). Se um artefato para o seu projeto não estiver listado em Versões do componente, especifique a versão que está listada para o Hive e o Hadoop nessa versão. Por exemplo, para componentes do Hadoop na versão 5.18.0 do Amazon EMR, a versão é `2.8.4-amzn-1`.

O URL do repositório de artefatos tem a seguinte sintaxe:

```
https://s3-endpoint/region-ID-emr-artifacts/emr-release-label/repos/maven/
```

- `s3-endpoint` é o endpoint do Amazon Simple Storage Service (Amazon S3) da região para o repositório, e `region-ID` é a região correspondente. Por exemplo, `s3.us-west-1.amazonaws.com` e `us-west-1`. Para obter mais informações, consulte os endpoints do Amazon S3 na Referência geral da Amazon Web Services. Não há diferença nos artefatos entre as regiões, portanto, você pode especificar a região mais conveniente para o seu ambiente de desenvolvimento.
- `emr-release-label` é a etiqueta de lançamento do cluster Amazon EMR que executará seu código. Os rótulos de versão estão no formato `emr-x.x.x`, como `emr-5.36.2`. Uma série de versões do EMR pode incluir várias versões. Por exemplo, se você estiver usando a versão 5.24.1 do EMR, use o primeiro rótulo de versão do EMR na série 5.24, `emr-5.24.0`, no URL do repositório de artefatos:

```
https://s3-endpoint/region-ID-emr-artifacts/emr-5.24.0/repos/maven/
```

Exemplo Configuração para pom.xml do Maven

O exemplo pom.xml a seguir configura um projeto Maven para compilar os artefatos do Apache Hadoop e do Apache Hive do emr-5.18.0 usando o repositório de artefatos em us-west-1. As versões de snapshot não estão disponíveis no repositório de artefatos. Portanto, os snapshots estão desabilitados no pom.xml. As reticências (...) no exemplo abaixo indicam omissão de outros parâmetros de configuração. Não os copie em seu projeto 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>
```

```
...  
</dependencies>  
  
</project>
```

EMR Sistema de arquivos (EMRFS)

O Sistema de EMR Arquivos (EMRFS) é uma implementação HDFS que todos os EMR clusters da Amazon usam para ler e gravar arquivos regulares da Amazon EMR diretamente no Amazon S3. EMRFS fornece a conveniência de armazenar dados persistentes no Amazon S3 para uso com o Hadoop, além de fornecer recursos como criptografia de dados.

A criptografia de dados permite criptografar objetos que EMRFS gravam no Amazon S3 e EMRFS permite trabalhar com objetos criptografados no Amazon S3. Se você estiver usando a EMR versão 4.8.0 ou posterior da Amazon, você pode usar configurações de segurança para configurar a criptografia para EMRFS objetos no Amazon S3, junto com outras configurações de criptografia. Para obter mais informações, consulte [Encryption options](#). Se você usa uma versão anterior da Amazon EMR, você pode definir manualmente as configurações de criptografia. Para ter mais informações, consulte [Especificando a criptografia do Amazon S3 usando propriedades EMRFS](#).

O Amazon S3 oferece uma forte consistência de leitura após gravação para todos GET e LIST operações em todos. PUT Regiões da AWS Isso significa que o que você escreve usando EMRFS é o que você lerá do Amazon S3, sem impacto no desempenho. Para obter mais informações, consulte [Modelo de consistência de dados do Amazon S3](#).

Ao usar a EMR versão 5.10.0 ou posterior da Amazon, você pode usar IAM funções diferentes para EMRFS solicitações ao Amazon S3 com base nos usuários do cluster, nos grupos ou na localização EMRFS dos dados no Amazon S3. Para obter mais informações, consulte [Configurar IAM funções para EMRFS solicitações ao Amazon S3](#).

Warning

Antes de ativar a execução especulativa para EMR clusters da Amazon que executam trabalhos do Apache Spark, consulte as informações a seguir.

EMRFS inclui o committer EMRFS otimizado para S3, uma OutputCommitter implementação otimizada para gravar arquivos no Amazon S3 durante o uso. EMRFS [Se você ativar o recurso de execução especulativa do Apache Spark com aplicativos que gravam dados no Amazon S3 e não usam o committer EMRFS otimizado para S3, poderá encontrar problemas de correção de dados descritos em -10063. SPARK](#) Isso pode ocorrer se você estiver usando EMR versões da Amazon anteriores EMR à versão 5.19 da Amazon ou se estiver gravando arquivos no Amazon S3 com formatos como e. ORC CSV Esses formatos não são compatíveis com o committer EMRFS otimizado para S3. Para obter uma lista completa

dos requisitos para usar o committer otimizado para EMRFS S3, consulte [Requisitos para o committer otimizado para S3](#). EMRFS

EMRFSa gravação direta geralmente é usada quando o committer EMRFS otimizado para S3 não é suportado, como ao escrever o seguinte:

- Um formato de saída diferente do Parquet, como ORC ou texto.
- Arquivos do Hadoop usando o Spark. RDD API
- Parquet usando Hive SerDe. Consulte [Hive metastore Parquet table conversion](#).

EMRFSa gravação direta não é usada nos seguintes cenários:

- Quando o committer EMRFS otimizado para S3 está ativado. Consulte [os requisitos para o committer EMRFS otimizado para S3](#).
- Ao escrever partições dinâmicas com `partitionOverwriteMode` definido como dinâmico.
- Ao gravar em locais de partição personalizados, como locais fora de conformidade com a convenção de localização de partições padrão do Hive.
- Ao usar sistemas de arquivos diferentes EMRFS, como gravar HDFS ou usar o sistema de arquivos S3A.

Para determinar se seu aplicativo usa gravação direta no Amazon EMR 5.14.0 ou posterior, habilite o registro no INFO Spark. Se uma linha de registro contendo o texto “Gravação direta:ENABLED” estiver presente nos registros do driver do Spark ou nos registros do contêiner do executor do Spark, seu aplicativo Spark gravou usando gravação direta. Por padrão, a execução especulativa é ativada OFF na AmazonEMRclusters. É altamente recomendável não ativar a execução especulativa se estas duas condições forem verdadeiras:

- Você está gravando dados no Amazon S3.
- Os dados são gravados em um formato diferente do Apache Parquet ou no formato Apache Parquet sem usar o committer otimizado para S3. EMRFS

Se você ativar a execução especulativa do Spark e gravar dados no Amazon S3 EMRFS usando gravação direta, poderá ocorrer perda intermitente de dados. Quando você grava ou HDFS grava dados no Parquet usando o committer EMRFS otimizado para S3, a EMR Amazon não usa gravação direta e esse problema não ocorre.

Se você precisar gravar dados em formatos que usam gravação EMRFS direta do Spark no Amazon S3 e usar execução especulativa, recomendamos gravar e transferir arquivos de saída HDFS para o Amazon S3 usando o S3DistCP.

Tópicos

- [Consistent view](#)
- [Autorizando o acesso aos EMRFS dados no Amazon S3](#)
- [Gerenciando o AWS Security Token Service endpoint padrão](#)
- [Especificando a criptografia do Amazon S3 usando propriedades EMRFS](#)

Consistent view

Warning

Em 1º de junho de 2023, a visualização EMRFS consistente alcançará o fim do suporte padrão para futuros EMR lançamentos da Amazon. EMRFSa visão consistente continuará funcionando para as versões existentes.

Com o lançamento da read-after-write consistência forte do Amazon S3 em 1º de dezembro de 2020, você não precisa mais usar a visualização EMRFS consistente (EMRFSCV) com seus clusters da AmazonEMR. EMRFSO CV é um recurso opcional que permite que os EMR clusters da Amazon verifiquem a lista e a read-after-write consistência dos objetos do Amazon S3. Quando você cria um cluster e o EMRFS CV é ativado, a Amazon EMR cria um banco de dados do Amazon DynamoDB para armazenar metadados de objetos que são usados para rastrear a lista read-after-write e a consistência dos objetos do S3. Agora você pode desativar o EMRFS CV e excluir o banco de dados do DynamoDB que ele usa para não acumular custos adicionais. Os procedimentos a seguir explicam como verificar o atributo CV, desativá-lo e excluir o banco de dados do DynamoDB usado pelo atributo.

Para verificar se você está usando o recurso EMRFS CV

1. Navegue até a guia Configuração. Se seu cluster tiver a configuração a seguir, ele usará EMRFS CV.

```
Classification=emrfs-site,Property=fs.s3.consistent,Value=true
```

2. Como alternativa, use o AWS CLI para descrever seu cluster com [describe-clusterAPI](#). Se a saída contiver `fs.s3.consistent: true`, seu cluster usará EMRFS CV.

Para desativar o EMRFS CV em seus EMR clusters da Amazon

Para desativar o recurso de EMRFS currículo, use uma das três opções a seguir. É necessário testar essas opções em seu ambiente de teste antes de aplicá-las aos ambientes de produção.

1. Para interromper seu cluster existente e iniciar um novo cluster sem opções de EMRFS CV.
 - a. Antes de interromper o cluster, faça backup dos dados e notifique os usuários.
 - b. Para interromper o cluster, siga as instruções apresentadas em [Terminate a cluster](#).
 - c. Se você usa o EMR console da Amazon para criar um novo cluster, navegue até Opções avançadas. Na seção Editar configurações do software, desmarque a opção de ativar o EMRFS CV. Se a caixa de seleção para exibição EMRFS consistente estiver disponível, mantenha-a desmarcada.
 - d. Se você costuma AWS CLI criar um novo cluster com o [create-clusterAPI](#), não use a `--emrfs` opção, que ativa o EMRFS CV.
 - e. Se você usar um SDK ou AWS CloudFormation para criar um novo cluster, não use nenhuma das configurações listadas em [Configurar visualização consistente](#).
2. Para clonar um cluster e remover EMRFS o CV
 - a. No EMR console da Amazon, escolha o cluster que usa o EMRFS CV.
 - b. Na parte superior da página Detalhes do cluster, escolha Clonar.
 - c. Escolha Anterior e navegue até a Etapa 1: software e etapas.
 - d. Em Editar configurações do software, remova o EMRFS CV. Em Editar configuração, exclua as seguintes configurações na classificação `emrfs-site`. Se você estiver carregando JSON de um bucket do S3, deverá modificar seu objeto do 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. Para remover o EMRFS CV de um cluster que usa grupos de instâncias
 - a. Use o comando a seguir para verificar se um único EMR cluster usa a tabela do DynamoDB EMRFS associada ao CV ou se vários clusters compartilham a tabela. O nome da tabela é especificado em `fs.s3.consistent.metadata.tableName`, conforme descrito em [Configure consistent view](#). O nome da tabela padrão usado pelo EMRFS CV é `EmrFSMetadata`.

```

aws emr describe-cluster --cluster-id j-XXXXX | grep
fs.s3.consistent.metadata.tableName

```

- b. Se seu cluster não compartilhar seu banco de dados do DynamoDB com outro cluster, use o comando a seguir para reconfigurar o cluster e desativar o CV. EMRFS Para obter mais informações, consulte [Reconfigure an instance group in a running cluster](#).

```

aws emr modify-instance-groups --cli-input-json file://disable-emrfs-1.json

```

Esse comando abrirá o arquivo que você deseja modificar. Modifique o arquivo com as configurações a seguir.

```

{
  "ClusterId": "j-xxxx",
  "InstanceGroups": [
    {
      "InstanceGroupId": "ig-xxxx",
      "Configurations": [
        {
          "Classification": "emrfs-site",
          "Properties": {
            "fs.s3.consistent": "false"
          }
        },
        "Configurations": []
      ]
    }
  ]
}

```

```

    }
  ]
}
```

- c. Se seu cluster compartilhar a tabela do DynamoDB com outro cluster, EMRFS desative o CV em todos os clusters no momento em que nenhum cluster modifique nenhum objeto no local compartilhado do S3.

Para excluir recursos do Amazon DynamoDB associados ao CV EMRFS

Depois de remover o EMRFS CV dos seus EMR clusters da Amazon, exclua os recursos do DynamoDB associados ao CV. EMRFS Até fazer isso, você continuará incorrendo nas cobranças do DynamoDB associadas ao CV. EMRFS

1. Verifique as CloudWatch métricas da sua tabela do DynamoDB e confirme se a tabela não é usada por nenhum cluster.
2. Exclua a tabela do DynamoDB.

```
aws dynamodb delete-table --table-name <your-table-name>
```

Para excluir SQS recursos da Amazon associados ao EMRFS CV

1. Se você configurou seu cluster para enviar notificações de inconsistência para a AmazonSQS, você pode excluir todas as SQS filas.
2. Encontre o nome da SQS fila da Amazon especificado `emfs.s3.consistent.notification.SQS.queueName`, conforme descrito em [Configurar visualização consistente](#). O formato padrão do nome da fila é `EMRFS-Inconsistency-<j-cluster ID>`.

```
aws sqs list-queues | grep 'EMRFS-Inconsistency'
aws sqs delete-queue --queue-url <your-queue-url>
```

Para parar de usar o EMRFS CLI

- O [EMRFSCLI](#) gerencia os metadados que o EMRFS CV gera. À medida que o suporte padrão para o EMRFS CV chegar ao fim em futuros lançamentos da AmazonEMR, o suporte para o também EMRFS CLI chegará ao fim.

Tópicos

- [Habilitar a visualização consistente](#)
- [Entendendo como a visualização EMRFS consistente rastreia objetos no Amazon S3](#)
- [Lógica de novas tentativas](#)
- [EMRFSmetadados de visualização consistente](#)
- [Configure notificações de consistência para CloudWatch a Amazon SQS](#)
- [Configurar a visualização consistente](#)
- [EMRFSCLIReferência de comando](#)

Habilitar a visualização consistente

Você pode habilitar a criptografia do lado do servidor do Amazon S3 ou a visualização consistente para EMRFS usar o AWS Management Console, AWS CLI, ou a classificação de configuração.

`emrfs-site`


Para configurar a exibição consistente usando o console

1. Navegue até o novo EMR console da Amazon e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte [Usar o console antigo](#).
2. Escolha Create cluster (Criar cluster), Go to advanced options (Ir para opções avançadas).
3. Escolha as configurações de Step 1: Software and Steps (Etapa 1: Software e etapas) e Step 2: Hardware (Etapa 2: Hardware).
4. Para a Etapa 3: Configurações gerais do cluster, em Opções adicionais, escolha uma exibição EMRFS consistente.
5. Em EMRFSArmazenamento de metadados, digite o nome do seu armazenamento de metadados. O valor padrão é **EmrFSMetadata**. Se a mrFSMetadata tabela E não existir, ela será criada para você no DynamoDB.

Note

EMRA Amazon não remove automaticamente os EMRFS metadados do DynamoDB quando o cluster é encerrado.


6. Para Number of retries (Número de novas tentativas), digite um valor inteiro. Se uma inconsistência for detectada, EMRFS tente ligar para o Amazon S3 esse número de vezes. O valor padrão é **5**.
7. Para Retry period (in seconds) [Período de repetição (em segundos)], digite um valor inteiro. Essa é a quantidade de tempo de EMRFS espera entre as tentativas de repetição. O valor padrão é **10**.

 Note

Tentativas subsequentes usam um recuo exponencial.

Para iniciar um cluster com visualização consistente ativada usando o AWS CLI

Recomendamos que você instale a versão atual do AWS CLI. Para baixar a versão mais recente, consulte <https://aws.amazon.com/cli/>.

 Note

Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

```
aws emr create-cluster --instance-type m5.xlarge --instance-count 3 --emrfs  
  Consistent=true \  
--release-label emr-7.2.0 --ec2-attributes KeyName=myKey
```

Para verificar se a visualização consistente está ativada usando o AWS Management Console

- Para verificar se a exibição consistente está habilitada no console, navegue até Cluster List (Lista de clusters) e selecione o nome do cluster para visualizar Cluster Details (Detalhes do cluster). O campo “visualização EMRFS consistente” tem um valor de Enabled ou Disabled.

Para verificar se a exibição consistente está habilitada, examinando o arquivo **emrfs-site.xml**

- Você pode verificar se a consistência está habilitada inspecionando o arquivo de configuração `emrfs-site.xml` no nó principal do cluster. Se o valor booleano para `fs.s3.consistent` estiver definido como `true`, a visualização consistente estará habilitada para operações do sistema de arquivo envolvendo o Amazon S3.

Entendendo como a visualização EMRFS consistente rastreia objetos no Amazon S3

EMRFS cria uma visão consistente dos objetos no Amazon S3 adicionando informações sobre esses objetos aos metadados. EMRFS adiciona essas listagens aos seus metadados quando:

- Um objeto escrito por EMRFS durante o curso de um EMR trabalho na Amazon.
- Um objeto é sincronizado ou importado para EMRFS metadados usando o EMRFS CLI

Os objetos lidos por não EMRFS são adicionados automaticamente aos metadados. Quando EMRFS exclui um objeto, uma listagem ainda permanece nos metadados com um estado excluído até que a listagem seja removida usando o EMRFS CLI. Para saber mais sobre o CLI, consulte [EMRFS CLI Referência de comando](#). Para obter mais informações sobre como limpar listagens nos EMRFS metadados, consulte [EMRFS metadados de visualização consistente](#)

Para cada operação do Amazon S3, EMRFS verifica os metadados em busca de informações sobre o conjunto de objetos em uma visualização consistente. Se EMRFS descobrir que o Amazon S3 é inconsistente durante uma dessas operações, ele repete a operação de acordo com os parâmetros definidos nas propriedades de configuração. `emrfs-site` Depois de EMRFS esgotar as novas tentativas, ele lança uma `ConsistencyException` ou registra a exceção e continua o fluxo de trabalho. Para obter mais informações sobre a lógica de novas tentativas, consulte [Lógica de novas tentativas](#). Você pode encontrar `ConsistencyExceptions` nos seus logs, por exemplo:

- `listStatus`: Nenhum objeto Amazon S3 para item de metadados `/S3_bucket/dir/object`
- `getFileStatus`: a chave `dir/file` está presente nos metadados, mas não no Amazon S3

Se você excluir um objeto diretamente do Amazon S3, essa visualização EMRFS consistente rastreia, EMRFS tratará esse objeto como inconsistente porque ele ainda está listado nos metadados como presente no Amazon S3. Se seus metadados ficarem fora de sincronia com as EMRFS faixas

de objetos no Amazon S3, você poderá usar `sync` o subcomando do para redefinir os metadados de forma que eles reflitam EMRFS CLI o Amazon S3. Para descobrir discrepâncias entre os metadados e o Amazon S3, use o `diff`. Por fim, EMRFS só tem uma visão consistente dos objetos referenciados nos metadados; pode haver outros objetos no mesmo caminho do Amazon S3 que não estão sendo rastreados. Quando EMRFS lista os objetos em um caminho do Amazon S3, ele retorna o superconjunto dos objetos que estão sendo rastreados nos metadados e aqueles nesse caminho do Amazon S3.

Lógica de novas tentativas

EMRFS tenta verificar a consistência da lista de objetos rastreados em seus metadados para um número específico de novas tentativas. O padrão é 5. No caso em que o número de novas tentativas for excedido, o trabalho originário retornará uma falha, a não ser que `fs.s3.consistent.throwExceptionOnInconsistency` seja definido como `false`, onde ele registrará somente os objetos controlados como inconsistentes. EMRFS usa uma política de repetição de recuo exponencial por padrão, mas você também pode defini-la como uma política fixa. Os usuários também podem querer fazer novas tentativas durante um determinado período antes de prosseguirem com o restante do seu trabalho sem lançar uma exceção. Eles podem conseguir isso definindo `fs.s3.consistent.throwExceptionOnInconsistency` como `false`, `fs.s3.consistent.retryPolicyType` como `fixed` e `fs.s3.consistent.retryPeriodSeconds` para o valor desejado. O exemplo a seguir cria um cluster com a consistência habilitada, o que registra as inconsistências e define um intervalo de repetição fixo de 10 segundos:

Exemplo Definir o período de repetição como um valor fixo

```
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

Os caracteres de continuação de linha do Linux (`\`) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (`^`).

Para ter mais informações, consulte [Consistent view](#).

EMRFSconfigurações para IMDS receber chamadas regionais

EMRFS depende do IMDS (serviço de metadados da instância) para obter a região da instância e o Amazon S3, o DynamoDB ou os endpoints. AWS KMS No entanto, IMDS tem um limite de quantas solicitações ele pode processar, e as solicitações que excederem esse limite falharão. Esse IMDS limite pode causar EMRFS falhas na inicialização e fazer com que a consulta ou o comando falhem. É possível usar o mecanismo de repetição de recuo exponencial aleatório a seguir e as propriedades de configuração de uma região de fallback em `emrfs-site.xml` para resolver o cenário em que todas as tentativas falham.

```
<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>
```

EMRFSmetadados de visualização consistente

EMRFS a visualização consistente rastreia a consistência usando uma tabela do DynamoDB para rastrear objetos no Amazon S3 que foram sincronizados ou criados por. EMRFS Os metadados são usados para controlar todas as operações (ler, gravar, atualizar e copiar), e nenhum conteúdo real é armazenado neles. Esses metadados são usados para validar se os objetos ou metadados

recebidos do Amazon S3 correspondem ao que se espera. Essa confirmação permite verificar EMRFS a consistência e read-after-write a consistência da lista para novas EMRFS gravações de objetos no Amazon S3 ou objetos sincronizados com. EMRFS Vários clusters podem compartilhar os mesmos metadados.

Como adicionar entradas aos metadados

Você pode usar os subcomandos `sync` ou `import` para adicionar entradas aos metadados. `sync` reflete o estado dos objetos do Amazon S3 em um caminho, enquanto `import` é usado estritamente para adicionar novas entradas aos metadados. Para ter mais informações, consulte [EMRFSCLIReferência de comando](#).

Como verificar diferenças entre metadados e objetos no Amazon S3

Para verificar as diferenças entre os metadados e o Amazon S3, use `diff` o subcomando do. EMRFS CLI Para ter mais informações, consulte [EMRFSCLIReferência de comando](#).

Como saber se as operações de metadados estão sendo limitadas

EMRFS define limites de capacidade de taxa de transferência padrão nos metadados para suas operações de leitura e gravação em 500 e 100 unidades, respectivamente. Grandes número de objetos ou buckets podem fazer com que as operações excedam essa capacidade e, nesse ponto, o DynamoDB limitará as operações. Por exemplo, um aplicativo pode causar EMRFS o lançamento de um `ProvisionedThroughputExceededException` se você executar uma operação que exceda esses limites de capacidade. Após a limitação, a EMRFS CLI ferramenta tenta gravar novamente na tabela do DynamoDB usando o [backoff exponencial](#) até que a operação seja concluída ou quando atingir o valor máximo de repetição para gravar objetos da Amazon no Amazon S3. EMR

É possível configurar seus próprios limites de capacidade de throughput. No entanto, o DynamoDB tem limites rígidos de partição de 3.000 unidades de capacidade de leitura RCUs () e 1.000 unidades de capacidade de gravação WCUs () por segundo para operações de leitura e gravação. Para evitar `sync` falhas causadas pela limitação, recomendamos que você limite a taxa de transferência das operações de leitura a menos de 3.000 RCUs e as operações de gravação a menos de 1.000. WCUs Para obter instruções sobre como definir limites personalizados de capacidade de throughput, consulte [Configurar a visualização consistente](#).

Você também pode ver CloudWatch as métricas da Amazon para seus EMRFS metadados no console do DynamoDB, onde você pode ver o número de solicitações limitadas de leitura e gravação. Se você tiver um valor diferente de zero para solicitações limitadas, seu aplicativo poderá se

beneficiar com o aumento da capacidade de throughput alocada para operações de leitura ou gravação. Você também poderá perceber um benefício de desempenho se constatar que as suas operações estão se aproximando da capacidade de throughput máxima alocada em leituras ou gravações por um período prolongado.

Características de produtividade para operações notáveis EMRFS

O padrão para operações de leitura e gravação é de 400 e 100 unidades de capacidade de throughput, respectivamente. As seguintes características de desempenho dão uma ideia de qual throughput é necessário para determinadas operações. Esses testes foram realizados usando um cluster m3.large de nó único. Todas as operações foram definidas em um único thread. O desempenho tem uma variação considerável com base em características específicas dos aplicativos e pode exigir experimentação para otimizar as operações do sistema de arquivos.

Operation	Média read-per-second	Média write-per-second
create (objeto)	26.79	6.70
delete (objeto)	10.79	10.79
delete (diretório contendo 1000 objetos)	21.79	338.40
getFileStatus(objeto)	34.70	0
getFileStatus(diretório)	19.96	0
listStatus(diretório contendo 1 objeto)	43.31	0
listStatus(diretório contendo 10 objetos)	44.34	0
listStatus(diretório contendo 100 objetos)	84.44	0
listStatus(diretório contendo 1.000 objetos)	308.81	0

Operation	Média read-per-second	Média write-per-second
listStatus(diretório contendo 10.000 objetos)	416.05	0
listStatus(diretório contendo 100.000 objetos)	823.56	0
listStatus(diretório contendo 1 milhão de objetos)	882.36	0
mkdir (contínuo por 120 segundos)	24.18	4.03
mkdir	12.59	0
rename (objeto)	19.53	4.88
rename (diretório contendo 1.000 objetos)	23.22	339.34

Para enviar uma etapa que limpa dados antigos do seu repositório de metadados

Os usuários podem querer remover entradas específicas nos metadados baseados no DynamoDB. Isso pode ajudar a reduzir os custos de armazenamento associados à tabela. Os usuários têm a capacidade de limpar manualmente ou programaticamente entradas específicas usando o subcomando. EMRFS CLI `delete` No entanto, se você excluir entradas dos metadados, EMRFS não fará mais nenhuma verificação de consistência.

A limpeza programática após a conclusão de um trabalho pode ser feita enviando uma etapa final ao seu cluster, que executa um comando no. EMRFS CLI Por exemplo, digite o seguinte comando para enviar uma etapa ao seu cluster para excluir todas as entradas com mais de dois dias.

```
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"
  ]
}
```

```
}
```

Use o StepId valor retornado para verificar o resultado da operação nos registros.

Configure notificações de consistência para CloudWatch a Amazon SQS

Você pode ativar CloudWatch métricas e SQS mensagens da Amazon EMRFS para eventuais problemas de consistência no Amazon S3.

CloudWatch

Quando CloudWatch as métricas são habilitadas, uma métrica chamada Inconsistência é enviada sempre que uma FileSystem API chamada falha devido à consistência eventual do Amazon S3.

Para visualizar CloudWatch métricas para eventuais problemas de consistência do Amazon S3

Para visualizar a métrica de inconsistência no CloudWatch console, selecione as EMRFS métricas e, em seguida, selecione um par JobFlowId/Nome da métrica. Por exemplo: j-162XXXXXXM2CU ListStatus, j-162XXXXXXM2CU GetFileStatus e assim por diante.

1. Abra o CloudWatch console em <https://console.aws.amazon.com/cloudwatch/>.
2. No Painel, na seção Métricas, escolha EMRFS.
3. No painel Job Flow Metrics, selecione um ou mais pares JobFlowId/Metric Name. Uma representação gráfica das métricas aparece na janela abaixo.

Amazon SQS

Quando SQS as notificações da Amazon são ativadas, uma SQS fila da Amazon com o nome EMRFS-Inconsistency-<jobFlowId> é criada quando EMRFS é inicializada. As SQS mensagens da Amazon são colocadas na fila quando uma FileSystem API chamada falha devido à eventual consistência do Amazon S3. A mensagem contém informações como JobFlowId,API, uma lista de caminhos inconsistentes, um rastreamento de pilha e assim por diante. As mensagens podem ser lidas usando o SQS console da Amazon ou usando o EMRFS `read-sqs` comando.

Para gerenciar SQS mensagens da Amazon para eventuais problemas de consistência do Amazon S3

SQSA's mensagens da Amazon para eventuais problemas de consistência do Amazon S3 podem ser lidas usando o. EMRFS CLI Para ler mensagens de uma SQS fila da EMRFS Amazon, digite o

`read-sqs` comando e especifique um local de saída no sistema de arquivos local do nó principal para o arquivo de saída resultante.

Você também pode excluir uma SQS fila EMRFS da Amazon usando o `delete-sqs` comando.

1. Para ler mensagens de uma SQS fila da Amazon, digite o seguinte comando. Substituir *queuename* com o nome da SQS fila da Amazon que você configurou e substituiu */path/filename* com o caminho para o arquivo de saída:

```
emrfs read-sqs --queue-name queuename --output-file /path/filename
```

Por exemplo, para ler e enviar SQS mensagens da Amazon da fila padrão, digite:

```
emrfs read-sqs --queue-name EMRFS-Inconsistency-j-162XXXXXXM2CU --output-file /path/filename
```

Note

Você também pode usar os atalhos `-q` e `-o` em vez de `--queue-name` e `--output-file`, respectivamente.

2. Para excluir uma SQS fila da Amazon, digite o seguinte comando:

```
emrfs delete-sqs --queue-name queuename
```

Por exemplo, para excluir a fila padrão, digite:

```
emrfs delete-sqs --queue-name EMRFS-Inconsistency-j-162XXXXXXM2CU
```

Note

Você também pode usar o atalho `-q` no lugar de `--queue-name`.

Configurar a visualização consistente

Você pode definir configurações adicionais para a visualização consistente usando propriedades de configuração para propriedades do `emrfs-site`. Por exemplo, você pode escolher uma taxa de

transferência padrão diferente do DynamoDB fornecendo os seguintes argumentos para a opção, usando CLI `--emrfs` a classificação de configuração `emrfs-site` (somente a versão 4.x e posterior da EMR Amazon) ou uma ação de bootstrap para configurar o arquivo `emrfs-site.xml` no nó principal:

Example Alterar os valores de leitura e gravação de metadados padrão na execução do 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
```

Como alternativa, use o seguinte arquivo de configuração e salve-o localmente ou no Amazon S3:

```
[  
  {  
    "Classification": "emrfs-site",  
    "Properties": {  
      "fs.s3.consistent.metadata.read.capacity": "600",  
      "fs.s3.consistent.metadata.write.capacity": "300"  
    }  
  }  
]
```

Use a configuração que você criou com a seguinte sintaxe:

```
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

Os caracteres de continuação de linha do Linux (`\`) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (`^`).

As opções a seguir podem ser definidas usando configurações ou AWS CLI `--emrfs` argumentos. Para obter mais informações sobre esses argumentos, consulte [AWS CLI Command Reference](#).

Propriedades `emrfs-site.xml` para a visualização consistente

Propriedade	Valor padrão	Descrição
<code>fs.s3.consistent</code>	false	Quando definida como true , essa propriedade é configurada EMRFS para usar o DynamoDB para fornecer consistência.
<code>fs.s3.consistent.retryPolicyType</code>	exponential	Essa propriedade identifica a política a ser usada em caso de novas tentativas para problemas de consistência. As opções incluem: <code>exponential</code> , <code>fixed</code> ou <code>none</code> .
<code>fs.s3.consistent.retryPeriodSeconds</code>	1	Essa propriedade define o tempo de espera entre tentativas de repetição de consistência.
<code>fs.s3.consistent.retryCount</code>	10	Essa propriedade define o número máximo de tentativas quando é detectada uma inconsistência.
<code>fs.s3.consistent.throwExceptionOnInconsistency</code>	true	Essa propriedade determina se uma exceção de consistência deve ser lançada ou registrada. Quando definida como true , um <code>ConsistencyException</code> é lançado.
<code>fs.s3.consistent.metadata.autoCreate</code>	true	Quando definida como true , essa propriedade permite a criação automática de tabelas de metadados.

Propriedade	Valor padrão	Descrição
<code>fs.s3.consistent.metadata.tag verification.enabled</code>	true	Com o Amazon EMR 5.29.0, essa propriedade é ativada por padrão. Quando ativado, EMRFS usa o S3 ETags para verificar se os objetos que estão sendo lidos são da versão mais recente disponível. Esse recurso é útil para casos de read-after-update uso em que os arquivos no S3 estão sendo sobrescritos, mantendo o mesmo nome. Atualmente, esse recurso de ETag verificação não funciona com o S3 Select.
<code>fs.s3.consistent.metadata.tableName</code>	EmrFSMetadata	Essa propriedade especifica o nome da tabela de metadados no DynamoDB.
<code>fs.s3.consistent.metadata.read.capacity</code>	500	Essa propriedade especifica a capacidade de leitura do DynamoDB a ser provisionada quando a tabela de metadados é criada.
<code>fs.s3.consistent.metadata.write.capacity</code>	100	Essa propriedade especifica a capacidade de gravação do DynamoDB a ser provisionada quando a tabela de metadados é criada.

Propriedade	Valor padrão	Descrição
<code>fs.s3.consistent.fastList</code>	true	Quando definida como true , essa propriedade usa vários threads para listar um diretório (quando necessário). A consistência deve ser habilitada para usar essa propriedade.
<code>fs.s3.consistent.fastList.p refetchMetadata</code>	false	Quando definida como true , essa propriedade permite que metadados façam uma pré-busca por diretórios que contêm mais de 20.000 itens.
<code>fs.s3.consistent.notification.CloudWatch</code>	false	Quando definidas como true , as CloudWatch métricas são habilitadas para FileSystem API chamadas que falham devido a eventuais problemas de consistência do Amazon S3.
<code>fs.s3.consistent.notification.SQS</code>	false	Quando definido como true , eventuais notificações de consistência são enviadas para uma SQS fila da Amazon.
<code>fs.s3.consistent.notification.SQS.queueName</code>	EMRFS-Inc onsistenc y- <jobFlo wId>	A alteração dessa propriedade permite que você especifique seu próprio nome de SQS fila para mensagens relacionadas a eventuais problemas de consistência do Amazon S3.

Propriedade	Valor padrão	Descrição
<code>fs.s3.consistent.notification.SQS.customMsg</code>	none	Essa propriedade permite que você especifique informações personalizadas incluídas nas SQS mensagens sobre eventuais problemas de consistência do Amazon S3. Se um valor não for especificado para essa propriedade, o campo correspondente na mensagem ficará vazio.
<code>fs.s3.consistent.dynamodb.endpoint</code>	none	Essa propriedade permite especificar um endpoint personalizado do DynamoDB para os seus metadados de exibição consistente.
<code>fs.s3.useRequesterPaysHeader</code>	false	Quando definida como true , essa propriedade permite solicitações do Amazon S3 para buckets com a opção de pagante de solicitações habilitada.

EMRFSCLIReferência de comando

O EMRFS CLI é instalado por padrão em todos os nós principais do cluster criados usando a EMR versão 3.2.1 ou posterior da Amazon. Você pode usar o EMRFS CLI para gerenciar os metadados para uma visualização consistente.

Note

O `emrfs` comando só é suportado com a emulação do terminal VT100. No entanto, ele pode funcionar com outros modos de emulador de terminal.

comando de nível superior emrfs

O comando de nível superior emrfs oferece suporte à seguinte estrutura.

```
emrfs [describe-metadata | set-metadata-capacity | delete-metadata | create-metadata |
 \
 list-metadata-stores | diff | delete | sync | import ] [options] [arguments]
```

Especifique [opções], com ou sem [argumentos] conforme descrito na tabela a seguir. Para [opções] específicas de subcomandos (describe-metadata, set-metadata-capacity, etc.), consulte cada subcomando abaixo.

[Opções] para emrfs

Opção	Descrição	Obrigatório
-a <i>AWS_ACCESS_KEY_ID</i> --access-key <i>AWS_ACCESS_KEY_ID</i>	A chave de AWS acesso que você usa para gravar objetos no Amazon S3 e para criar ou acessar um armazenamento de metadados no DynamoDB. Por padrão, <i>AWS_ACCESS_KEY_ID</i> é definido como a chave de acesso usada para criar o cluster.	Não
-s <i>AWS_SECRET_ACCESS_KEY</i> --secret-key <i>AWS_SECRET_ACCESS_KEY</i>	A chave AWS secreta associada à chave de acesso que você usa para gravar objetos no Amazon S3 e para criar ou acessar um armazenamento de metadados no DynamoDB. Por padrão, <i>AWS_SECRET_ACCESS_KEY</i> é definido como a chave secreta associada à chave de acesso usada para criar o cluster.	Não
-v --verbose	Torna a saída detalhada.	Não
-h --help	Exibe a mensagem de ajuda para o comando emrfs com uma instrução de uso.	Não

subcomando emrfs describe-metadata

[Opções] para emrfs describe-metadata

Opção	Descrição	Obrigatório
-m <i>METADATA_NAME</i> --metadata-name <i>METADATA_NAME</i>	<i>METADATA_NAME</i> é o nome da tabela de metadados do DynamoDB. Se o <i>METADATA_NAME</i> o argumento não é fornecido, o valor padrão é EmrFSMetadata .	Não

Exemplo exemplo de emrfs describe-metadata

O exemplo a seguir descreve a tabela de metadados padrão.

```
$ emrfs describe-metadata
EmrFSMetadata
  read-capacity: 400
  write-capacity: 100
  status: ACTIVE
  approximate-item-count (6 hour delay): 12
```

subcomando emrfs set-metadata-capacity

[Opções] para emrfs set-metadata-capacity

Opção	Descrição	Obrigatório
-m <i>METADATA_NAME</i> --metadata-name <i>METADATA_NAME</i>	<i>METADATA_NAME</i> é o nome da tabela de metadados do DynamoDB. Se o <i>METADATA_NAME</i> o argumento não é fornecido, o valor padrão é EmrFSMetadata .	Não
	A capacidade de throughput de leitura solicitada para a tabela de metadados. Se o <i>READ_CAPA</i>	Não

Opção	Descrição	Obrigatório
<code>-r <i>READ_CAPACITY</i> --read-capacity <i>READ_CAPACITY</i></code>	<i>CITY</i> o argumento não é fornecido, o valor padrão é 400.	
<code>-w <i>WRITE_CAPACITY</i> --write-capacity <i>WRITE_CAPACITY</i></code>	A capacidade de throughput de gravação solicitada para a tabela de metadados. Se o <i>WRITE_CAPACITY</i> o argumento não é fornecido, o valor padrão é 100.	Não

Exemplo exemplo de emrfs set-metadata-capacity

O exemplo a seguir define a capacidade de throughput de leitura como 600 e a capacidade de gravação como 150 para uma tabela de metadados denominada 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
```

subcomando emrfs delete-metadata

[Opções] para emrfs delete-metadata

Opção	Descrição	Obrigatório
<code>-m <i>METADATA_NAME</i> --metadata-name <i>METADATA_NAME</i></code>	<i>METADATA_NAME</i> é o nome da tabela de metadados do DynamoDB. Se o <i>METADATA_NAME</i> o argumento não é fornecido, o valor padrão é EmrFSMetadata .	Não

Example exemplo de delete-metadata

O exemplo a seguir exclui a tabela de metadados padrão.

```
$ emrfs delete-metadata
```

subcomando emrfs create-metadata

[Opções] para emrfs create-metadata

Opção	Descrição	Obrigatório
<code>-m <i>METADATA_NAME</i></code> <code>--metadata-name <i>METADATA_NAME</i></code>	<i>METADATA_NAME</i> é o nome da tabela de metadados do DynamoDB. Se o <i>METADATA_NAME</i> o argumento não é fornecido, o valor padrão é <code>EmrFSMetadata</code> .	Não
<code>-r <i>READ_CAPACITY</i></code> <code>--read-capacity <i>READ_CAPACITY</i></code>	A capacidade de throughput de leitura solicitada para a tabela de metadados. Se o <i>READ_CAPACITY</i> o argumento não é fornecido, o valor padrão é <code>400</code> .	Não
<code>-w <i>WRITE_CAPACITY</i></code> <code>--write-capacity <i>WRITE_CAPACITY</i></code>	A capacidade de throughput de gravação solicitada para a tabela de metadados. Se o <i>WRITE_CAPACITY</i> o argumento não é fornecido, o valor padrão é <code>100</code> .	Não

Example exemplo de emrfs create-metadata

O exemplo a seguir cria uma tabela de metadados chamada de `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
```

subcomando emrfs list-metadata-stores

O subcomando emrfs list-metadata-stores não tem [opções].

Example ist-metadata-stores Exemplo L

O exemplo a seguir lista suas tabelas de metadados.

```
$ emrfs list-metadata-stores
EmrFSMetadata
```

subcomando emrfs diff

[Opções] para emrfs diff

Opção	Descrição	Obrigatório
<code>-m <i>METADATA_NAME</i></code> <code>--metadata-name <i>METADATA_NAME</i></code>	<i>METADATA_NAME</i> é o nome da tabela de metadados do DynamoDB. Se o <i>METADATA_NAME</i> o argumento não é fornecido, o valor padrão é <code>EmrFSMetadata</code> .	Não
<code><i>s3://s3Path</i></code>	O caminho para o bucket do Amazon S3, para a comparação com a tabela de metadados. Buckets são sincronizados recursivamente.	Sim

Example exemplo de emrfs diff

O exemplo a seguir compara a tabela de metadados padrão com um bucket do 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

```

subcomando emrfs delete

[Opções] para emrfs delete

Opção	Descrição	Obrigatório
<code>-m <i>METADATA_NAME</i> --metadata-name <i>METADATA_NAME</i></code>	<i>METADATA_NAME</i> é o nome da tabela de metadados do DynamoDB. Se o <i>METADATA_NAME</i> o argumento não é fornecido, o valor padrão é <code>EmrFSMetadata</code> .	Não
<code>s3://s3Path</code>	O caminho para o bucket do Amazon S3 que você está rastreando para uma visualização consistente. Buckets são sincronizados recursivamente.	Sim
<code>-t <i>TIME</i> --time <i>TIME</i></code>	O tempo de expiração (interpretado usando o argumento de unidade de tempo). Todas as entradas de metadados anteriores ao <i>TIME</i> os argumentos são excluídos do bucket especificado.	
<code>-u <i>UNIT</i> --time-unit <i>UNIT</i></code>	A medida usada para interpretar o argumento <code>TIME</code> (nanossegundos, microssegundos, milissegundos, segundos, minutos, horas ou dias). Se nenhum argumento for especificado, o valor padrão será <code>days</code> .	

Opção	Descrição	Obrigatório
<code>--read-consumption</code> <i>READ_CONSUMPTION</i>	A quantidade solicitada de throughput de leitura disponível utilizado para a operação delete. Se o <i>READ_CONSUMPTION</i> o argumento não foi especificado, o valor padrão é 400.	Não
<code>--write-consumption</code> <i>WRITE_CONSUMPTION</i>	A quantidade solicitada de throughput de gravação disponível utilizado para a operação delete. Se o <i>WRITE_CONSUMPTION</i> o argumento não foi especificado, o valor padrão é 100.	Não

Exemplo exemplo de emrfs delete

O exemplo a seguir remove todos os objetos em um bucket do Amazon S3 dos metadados de rastreamento para uma visualização consistente.

```
$ emrfs delete s3://elasticmapreduce/samples/cloudfront
entries deleted: 11
```

subcomando emrfs import

[Opções] para emrfs import

Opção	Descrição	Obrigatório
<code>-m</code> <i>METADATA_NAME</i> <code>--metadata-name</code> <i>METADATA_NAME</i>	<i>METADATA_NAME</i> é o nome da tabela de metadados do DynamoDB. Se o <i>METADATA_NAME</i> o argumento não é fornecido, o valor padrão é <code>EmrFSMetadata</code> .	Não
<i>s3://s3Path</i>		Sim

Opção	Descrição	Obrigatório
	O caminho para o bucket do Amazon S3 que você está rastreando para uma visualização consistente. Buckets são sincronizados recursivamente.	
<code>--read-consumption</code> <i>READ_CONSUMPTION</i>	A quantidade solicitada de throughput de leitura disponível utilizado para a operação delete. Se o <i>READ_CONSUMPTION</i> o argumento não foi especificado, o valor padrão é 400.	Não
<code>--write-consumption</code> <i>WRITE_CONSUMPTION</i>	A quantidade solicitada de throughput de gravação disponível utilizado para a operação delete. Se o <i>WRITE_CONSUMPTION</i> o argumento não foi especificado, o valor padrão é 100.	Não

Exemplo exemplo de emrfs import

O exemplo a seguir importa todos os objetos em um bucket do Amazon S3 com os metadados de rastreamento para uma visualização consistente. Todas as chaves desconhecidas são ignoradas.

```
$ emrfs import s3://elasticmapreduce/samples/cloudfront
```

subcomando emrfs sync

[Opções] para emrfs sync

Opção	Descrição	Obrigatório
<code>-m</code> <i>METADATA_NAME</i> <code>--metadata-name</code> <i>METADATA_NAME</i>	<i>METADATA_NAME</i> é o nome da tabela de metadados do DynamoDB. Se o <i>METADATA_NAME</i> o argumento não é fornecido, o valor padrão é <code>EmrFSMetadata</code> .	Não

Opção	Descrição	Obrigatório
<i>s3://s3Path</i>	O caminho para o bucket do Amazon S3 que você está rastreando para uma visualização consistente. Buckets são sincronizados recursivamente.	Sim
<code>--read-consumption</code> <i>READ_CONSUMPTION</i>	A quantidade solicitada de throughput de leitura disponível utilizado para a operação delete. Se o <i>READ_CONSUMPTION</i> o argumento não foi especificado, o valor padrão é400.	Não
<code>--write-consumption</code> <i>WRITE_CONSUMPTION</i>	A quantidade solicitada de throughput de gravação disponível utilizado para a operação delete. Se o <i>WRITE_CONSUMPTION</i> o argumento não foi especificado, o valor padrão é100.	Não

Example exemplo do comando emrfs sync

O exemplo a seguir importa todos os objetos em um bucket do Amazon S3 com os metadados de rastreamento para uma visualização consistente. Todas as chaves desconhecidas são excluídas.

```
$ 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
```

subcomando emrfs read-sqs

[Opções] para emrfs read-sqs

Opção	Descrição	Obrigatório
<code>-q <i>QUEUE_NAME</i> --queue-name <i>QUEUE_NAME</i></code>	<i>QUEUE_NAME</i> é o nome da SQS fila da Amazon configurada em <code>emrfs-site.xml</code> . O valor padrão é EMRFS-Inconsistency-<jobFlowId> .	Sim
<code>-o <i>OUTPUT_FILE</i> --output-file <i>OUTPUT_FILE</i></code>	<i>OUTPUT_FILE</i> é o caminho para o arquivo de saída no sistema de arquivos local do nó principal. Mensagens lidas na fila são gravadas nesse arquivo.	Sim

subcomando emrfs delete-sqs

[Opções] para emrfs delete-sqs

Opção	Descrição	Obrigatório
<code>-q <i>QUEUE_NAME</i> --queue-name <i>QUEUE_NAME</i></code>	<i>QUEUE_NAME</i> é o nome da SQS fila da Amazon configurada em <code>emrfs-site.xml</code> . O valor padrão é EMRFS-Inconsistency-<jobFlowId> .	Sim

Envio de EMRFS CLI comandos como etapas

O exemplo a seguir mostra como usar o `emrfs` utilitário no nó principal aproveitando o AWS CLI ou API e o `command-runner.jar` para executar o `emrfs` comando como uma etapa. O exemplo usa o AWS SDK for Python (Boto3) para adicionar uma etapa a um cluster que adiciona objetos em um bucket do Amazon S3 à tabela de EMRFS metadados padrão.

```
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()
```

Você pode usar o valor de `step_id` retornado para verificar os logs em busca do resultado da operação.

Autorizando o acesso aos EMRFS dados no Amazon S3

Por padrão, a EMR função de EC2 determina as permissões para acessar EMRFS dados no Amazon S3. As IAM políticas anexadas a essa função se aplicam independentemente do usuário ou grupo que faz a solicitação EMRFS. O padrão é `EMR_EC2_DefaultRole`. Para obter mais informações, consulte [Função de serviço para EC2 instâncias de cluster \(perfil de EC2 instância\)](#).

A partir da EMR versão 5.10.0 da Amazon, você pode usar uma configuração de segurança para especificar IAM funções para EMRFS. Isso permite que você personalize permissões para EMRFS solicitações ao Amazon S3 para clusters que tenham vários usuários. Você pode especificar IAM funções diferentes para diferentes usuários e grupos e para diferentes locais de bucket do Amazon S3 com base no prefixo no Amazon S3. Quando EMRFS faz uma solicitação ao Amazon S3 que corresponde aos usuários, grupos ou locais que você especifica, o cluster usa a função correspondente que você especifica em vez da função para a EMR qual você especifica. Para obter mais informações, consulte [Configurar IAM funções para EMRFS solicitações ao Amazon S3](#).

Como alternativa, se sua EMR solução da Amazon tiver demandas além das IAM funções que EMRFS ela fornece, você pode definir uma classe de provedor de credenciais personalizada, que permite personalizar o acesso aos EMRFS dados no Amazon S3.

Criação de um provedor de credenciais personalizado para EMRFS dados no Amazon S3

[Para criar um provedor de credenciais personalizado, você implementa as classes configuráveis `AWSCredentialsProvider` do Hadoop.](#)

Para obter uma explicação detalhada dessa abordagem, consulte [Analisar com segurança os dados de outra AWS conta EMRFS no blog](#) de AWS Big Data. A postagem do blog inclui um tutorial que orienta você no processo end-to-end, desde a criação de IAM funções até o lançamento do cluster. Ela também fornece um exemplo de código Java que implementa a classe de provedor de credenciais personalizado.

As etapas básicas são:

Para especificar um provedor de credenciais personalizado

1. Crie uma classe de provedor de credenciais personalizada compilada como um JAR arquivo.
2. Execute um script como uma ação de bootstrap para copiar o JAR arquivo personalizado do provedor de credenciais para o `/usr/share/aws/emr/emrfs/auxlib` local no nó principal do cluster. Para obter mais informações sobre ações de bootstrap, consulte [\(Optional\) Create bootstrap actions to install additional software](#).
3. Personalize a `emrfs-site` classificação para especificar a classe que você implementa no JAR arquivo. Para obter mais informações sobre a especificação de objetos de configuração para personalizar aplicativos, consulte [Configuração de aplicativos](#) no Amazon EMR Release Guide.

O exemplo a seguir demonstra um comando `create-cluster` que executa um cluster do Hive com parâmetros de configuração comuns e também inclui:

- Uma ação de bootstrap que executa o script, `copy_jar_file.sh`, que é salvo no `mybucket` no Amazon S3.
- Uma `emrfs-site` classificação que especifica um provedor de credenciais personalizado definido no JAR arquivo como `MyCustomCredentialsProvider`

Note

Os caracteres de continuação de linha do Linux (`\`) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (`^`).

```
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":[]}]'
```

Gerenciando o AWS Security Token Service endpoint padrão

EMRFS usa o AWS Security Token Service (STS) para recuperar credenciais de segurança temporárias a fim de acessar seus AWS recursos. As EMR versões anteriores da Amazon enviam todas as AWS STS solicitações para um único endpoint global em `https://sts.amazonaws.com`. Em vez disso, as EMR versões 5.31.0 e 6.1.0 e posteriores da Amazon fazem solicitações para endpoints regionais AWS STS. Isso reduz a latência e melhora a validade do token da sessão. Para obter mais informações sobre AWS STS endpoints, consulte [Gerenciando AWS STS em uma AWS região](#) no Guia do AWS Identity and Access Management usuário.

Ao usar as versões 5.31.0 e 6.1.0 e posteriores da AmazonEMR, você pode substituir o endpoint padrão. Para isso, é necessário alterar a propriedade `fs.s3.sts.endpoint` da configuração `emrfs-site`.

O exemplo de AWS CLI a seguir define o AWS STS endpoint padrão usado por EMRFS para o endpoint 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

Os caracteres de continuação de linha do Linux (`\`) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (`^`).

Como alternativa, você pode criar um arquivo de JSON configuração usando o exemplo a seguir e especificá-lo usando o `--configurations` argumento de `emr create-cluster`. Para obter mais informações sobre o uso de `--configurations`, consulte [AWS CLI Command Reference](#).

```
[
{
```

```
"classification": "emrfs-site",
"properties": {
  "fs.s3.sts.endpoint": "https://sts.amazonaws.com"
}
}
```

Especificando a criptografia do Amazon S3 usando propriedades EMRFS

Important

A partir da EMR versão 4.8.0 da Amazon, você pode usar as configurações de segurança para aplicar as configurações de criptografia com mais facilidade e com mais opções. Recomendamos usar configuração de segurança. Para obter informações, consulte [Configure data encryption](#). As instruções do console descritas nesta seção estão disponíveis para versões anteriores à 4.8.0. Se você usar o AWS CLI para configurar a criptografia do Amazon S3 na configuração do cluster e em uma configuração de segurança nas versões subsequentes, a configuração de segurança substituirá a configuração do cluster.


Ao criar um cluster, você pode especificar criptografia do lado do servidor (SSE) ou criptografia do lado do cliente (CSE) para dados no EMRFS Amazon S3 usando o console ou usando propriedades de classificação por meio do `awscli`. `CSE emrfs-site` AWS CLI EMR SDK Amazon S3 SSE e Amazon CSE são mutuamente exclusivos; você pode escolher um deles, mas não os dois.

Para AWS CLI obter instruções, consulte a seção apropriada para seu tipo de criptografia abaixo.

Para especificar as opções de EMRFS criptografia usando o AWS Management Console

1. Navegue até o novo EMR console da Amazon e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte [Usar o console antigo](#).
2. Escolha Create cluster (Criar cluster), Go to advanced options (Ir para opções avançadas).
3. Escolha uma Release de 4.7.2 ou anterior.
4. Escolha outras opções para Software and Steps (Software e etapas), conforme apropriado para seu aplicativo e, em seguida, escolha Next (Próximo).

- Escolha as configurações nos painéis Hardware e General Cluster Settings (Configurações gerais do cluster), conforme apropriado para seu aplicativo.
- No painel Segurança, em Autenticação e criptografia, selecione a opção Criptografia S3 (comEMRFS) a ser usada.

 Note

A criptografia do lado do servidor S3 com gerenciamento de KMS chaves (SSE-KMS) não está disponível ao usar a EMR versão 4.4 ou anterior da Amazon.

- Se você escolher uma opção que usa o gerenciamento de AWS chaves, escolha uma ID de AWS KMS chave. Para ter mais informações, consulte [Usando AWS KMS keys para EMRFS criptografia](#).
 - Se você escolher a criptografia do lado do cliente S3 com o provedor de materiais personalizados, forneça o nome da classe e o local. JAR Para ter mais informações, consulte [Criptografia do lado do cliente do Amazon S3](#).
- Escolha outras opções conforme apropriado para seu aplicativo e, em seguida, escolha Create Cluster (Criar cluster).

Usando AWS KMS keys para EMRFS criptografia

A chave de AWS KMS criptografia deve ser criada na mesma região da sua instância de EMR cluster da Amazon e dos buckets do Amazon S3 usados com. EMRFS Se a chave especificada estiver em uma conta diferente daquela usada para configurar um cluster, você deverá especificar a chave usando suaARN.

A função do perfil da EC2 instância da Amazon deve ter permissões para usar a KMS chave que você especificar. A função padrão para o perfil da instância na Amazon EMR éEMR_EC2_DefaultRole. Se você usar uma função diferente para o perfil da instância ou usar IAM funções para EMRFS solicitações ao Amazon S3, certifique-se de que cada função seja adicionada como um usuário chave, conforme apropriado. Isso dá à função permissões para usar a KMS chave. Para obter mais informações, consulte [Usando políticas de chaves](#) no Guia do AWS Key Management Service desenvolvedor e [Configurar IAM funções para EMRFS solicitações ao Amazon S3](#).

Você pode usar o AWS Management Console para adicionar seu perfil de instância ou perfil de EC2 instância à lista de usuários de chave para a KMS chave especificada, ou você pode usar o AWS CLI ou an AWS SDK para anexar uma política de chaves apropriada.

Observe que a Amazon EMR oferece suporte somente a [KMSchaves simétricas](#). Você não pode usar uma [KMSchave assimétrica](#) para criptografar dados em repouso em um cluster da Amazon. Para obter ajuda para determinar se uma KMS chave é simétrica ou assimétrica, consulte [Identificação de chaves simétricas e assimétricas](#). KMS

O procedimento abaixo descreve como adicionar o perfil de EMR instância padrão da Amazon, `EMR_EC2_DefaultRole` como um usuário chave usando AWS Management Console o. Ele pressupõe que você já tenha criado uma KMS chave. Para criar uma nova KMS chave, consulte [Criação de chaves](#) no Guia do AWS Key Management Service desenvolvedor.

Para adicionar o perfil de EC2 instância da Amazon EMR à lista de usuários da chave de criptografia

1. Faça login no console AWS Management Console e abra o AWS Key Management Service (AWS KMS) em <https://console.aws.amazon.com/kms>.
2. Para alterar o Região da AWS, use o seletor de região no canto superior direito da página.
3. Selecione o alias da KMS chave a ser modificada.
4. Na página de detalhes da chave, em Key Users (Usuários de chaves), escolha Add (Adicionar).
5. Na caixa de diálogo Add key users (Adicionar usuários da chave) selecione a função apropriada. O nome da função padrão é `EMR_EC2_DefaultRole`.
6. Escolha Adicionar.

Criptografia do lado do servidor do Amazon S3

Quando você configura a criptografia do lado do servidor do Amazon S3, o Amazon S3 criptografa os dados no nível do objeto à medida que os grava no disco e os descriptografa quando são acessados. Para obter mais informações sobre issoSSE, consulte [Proteção de dados usando criptografia do lado do servidor no Guia do usuário](#) do Amazon Simple Storage Service.

Você pode escolher entre dois sistemas diferentes de gerenciamento de chaves ao especificar SSE na AmazonEMR:

- SSE-S3 — O Amazon S3 gerencia as chaves para você.

- SSE- KMS — Você usa um AWS KMS key para configurar políticas adequadas para a AmazonEMR. Para obter mais informações sobre os principais requisitos da AmazonEMR, consulte [Usando AWS KMS keys para criptografia](#).

SSEcom chaves fornecidas pelo cliente (SSE-C) não está disponível para uso com a Amazon. EMR

Para criar um cluster com SSE -S3 habilitado usando o AWS CLI

- Digite o seguinte comando:

```
aws emr create-cluster --release-label emr-4.7.2 or earlier \  
--instance-count 3 --instance-type m5.xlarge --emrfs Encryption=ServerSide
```

Você também pode ativar SSE -S3 configurando o `fs.s3.enableServerSidePropriedade` de criptografia para verdadeira nas `emrfs-site` propriedades. Veja o exemplo SSE KMS abaixo e omita a propriedade do ID da chave.

Para criar um cluster com SSE - KMS ativado usando o AWS CLI

Note

SSE- KMS está disponível somente na EMR versão 4.5.0 e posterior da Amazon.

- Digite o AWS CLI comando a seguir para criar um cluster com SSE -KMS, onde *keyID* é um AWS KMS key, por exemplo, *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]
```

--OU--

Digite o AWS CLI comando a seguir usando a `emrfs-site` classificação e forneça um JSON arquivo de configuração com conteúdo semelhante ao mostrado `myConfig.json` no exemplo abaixo:

```
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
```

Exemplo de conteúdo de myConfig.json:

```
[
  {
    "Classification": "emrfs-site",
    "Properties": {
      "fs.s3.enableServerSideEncryption": "true",
      "fs.s3.serverSideEncryption.kms.keyId": "a4567b8-9900-12ab-1234-123a45678901"
    }
  }
]
```

Propriedades de configuração para SSE -S3 e - SSE KMS

Essas propriedades podem ser configuradas usando a classificação de configuração `emrfs-site`. SSE- KMS está disponível somente na EMR versão 4.5.0 e posterior da Amazon.

Propriedade	Valor padrão	Descrição
<code>fs.s3.enableServerSideEncryption</code>	false	Quando definida como true , objetos armazenados no Amazon S3 são criptografados usando criptografia do lado do servidor. Se nenhuma chave for especificada, SSE -S3 será usado.
<code>fs.s3.serverSideEncryption.kms.keyId</code>	n/a	Especifica uma ID de AWS KMS chave para ARN. Se uma chave for especificada, SSE - KMS será usada.

Criptografia do lado do cliente do Amazon S3

Com a criptografia do lado do cliente do Amazon S3, a criptografia e a descriptografia do Amazon S3 ocorrem no cliente em seu cluster. EMRFS Os objetos são criptografados antes de serem carregados no Amazon S3 e descriptografados após serem baixados. O provedor especificado por você fornece a chave de criptografia que o cliente usa. O cliente pode usar chaves fornecidas por AWS KMS (CSE-KMS) ou uma classe Java personalizada que fornece a chave raiz do lado do cliente (CSE-C). As especificações da criptografia são ligeiramente diferentes entre CSE - KMS e CSE -C, dependendo do provedor especificado e dos metadados do objeto que está sendo descriptografado ou criptografado. Para obter mais informações sobre essas diferenças, consulte [Proteger dados usando a criptografia do lado do cliente](#) no Guia do usuário do Amazon Simple Storage Service.

Note

O Amazon S3 CSE só garante que EMRFS os dados trocados com o Amazon S3 sejam criptografados; nem todos os dados nos volumes de instâncias de cluster são criptografados. Além disso, como o Hue não usa EMRFS, os objetos que o Navegador de arquivos Hue S3 grava no Amazon S3 não são criptografados.

Para especificar CSE - KMS para EMRFS dados no Amazon S3 usando o AWS CLI

- Digite o seguinte comando e substitua *MyKMSKeyID* com o ID da chave ou ARN da KMS chave a ser usada:

```
aws emr create-cluster --release-label emr-4.7.2 or earlier
--emrfs Encryption=ClientSide,ProviderType=KMS,KMSKeyId=MyKMSKeyId
```

Criar um provedor de chaves personalizado

Dependendo do tipo de criptografia que você usa ao criar um provedor de chave personalizado, o aplicativo também deve implementar EncryptionMaterialsProvider interfaces diferentes. Ambas as interfaces estão disponíveis na versão 1.11.0 e posterior AWS SDK para Java.

- [Para implementar a criptografia do Amazon S3, use o com.amazonaws.services.s3.model.EncryptionMaterialsProvider](#) interface.

- Para implementar a criptografia de disco local, use [com.amazonaws.services.elasticmapreduce.spi.security. EncryptionMaterialsProvider](https://docs.aws.amazon.com/elasticmapreduce/spi/security/EncryptionMaterialsProvider) interface.

Você pode usar qualquer estratégia para fornecer materiais de criptografia para a implementação. Por exemplo, você pode optar por fornecer materiais de criptografia estática ou integrá-los a um sistema de gerenciamento de chaves mais complexo.

Se você estiver usando a criptografia do Amazon S3, deverá usar os algoritmos de criptografia AES/GCM/NoPadding para materiais de criptografia personalizados.

Se você estiver usando criptografia de disco local, o algoritmo de criptografia a ser usado para materiais de criptografia personalizados varia de acordo com a EMR versão. Para o Amazon EMR 7.0.0 e versões anteriores, você deve usar AES/GCM/NoPadding. Para Amazon EMR 7.1.0 e superior, você deve usar AES.

A `EncryptionMaterialsProvider` classe obtém materiais de criptografia por contexto de criptografia. A Amazon EMR preenche as informações do contexto de criptografia em tempo de execução para ajudar o chamador a determinar os materiais de criptografia corretos a serem devolvidos.

Example Exemplo: Usando um provedor de chave personalizado para criptografia do Amazon S3 com EMRFS

Quando a Amazon EMR busca os materiais de criptografia da `EncryptionMaterialsProvider` classe para realizar a criptografia, EMRFS opcionalmente preenche o `materialsDescription` argumento com dois campos: o Amazon S3 URI para o objeto e o `JobFlowId` do cluster, que podem ser usados pela `EncryptionMaterialsProvider` classe para retornar materiais de criptografia seletivamente.

Por exemplo, o provedor pode retornar chaves diferentes para diferentes prefixos do Amazon S3 URI. É a descrição dos materiais de criptografia retornados que são eventualmente armazenados com o objeto Amazon S3, em vez do `materialsDescription` valor que é gerado EMRFS e passado para o provedor. Ao descriptografar um objeto do Amazon S3, a descrição do material de criptografia é passada para a `EncryptionMaterialsProvider` classe, para que ela possa, novamente, retornar seletivamente a chave correspondente para descriptografar o objeto.

Uma implementação de `EncryptionMaterialsProvider` referência é fornecida abaixo. Outro provedor personalizado, [EMRFSRSAEncryptionMaterialsProvider](#), está disponível em GitHub.

```
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;
    }
}
```

```
}  
}
```

Especificando um fornecedor de materiais personalizados usando o AWS CLI

Para usar a AWS CLI, transmita os argumentos `Encryption`, `ProviderType`, `CustomProviderClass` e `CustomProviderLocation` para a opção `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
```

`Encryption` A configuração para `ClientSide` habilitar a criptografia do lado do cliente
`CustomProviderClass` é o nome do seu `EncryptionMaterialsProvider` objeto e
`CustomProviderLocation` é a localização local ou do Amazon S3 a partir da qual a EMR Amazon
`CustomProviderClass` copia para cada nó no cluster e o coloca no classpath.

Especificando um fornecedor de materiais personalizados usando um SDK

Para usar um SDK, você pode definir a propriedade `fs.s3.cse.encryptedMaterialsProvider.uri` para baixar a `EncryptionMaterialsProvider` classe personalizada que você armazena no Amazon S3 para cada nó do seu cluster. Você configura isso no `emrfs-site.xml` arquivo junto com a localização CSE ativada e correta do provedor personalizado.

Por exemplo, no AWS SDK for Java uso `RunJobFlowRequest`, seu código pode ter a seguinte aparência:

```
<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>
```

Personalizar EncryptionMaterialsProvider com argumentos

Pode ser necessário passar argumentos diretamente para o provedor. Para fazer isso, você pode usar a classificação de configuração `emrfs-site` com argumentos personalizados definidos como propriedades. Uma configuração de exemplo é mostrada a seguir, que é salva como um arquivo, `myConfig.json`:

```
[
  {
    "Classification": "emrfs-site",
    "Properties": {
      "myProvider.arg1": "value1",
      "myProvider.arg2": "value2"
    }
  }
]
```

Usando o `create-cluster` comando do AWS CLI, você pode usar a `--configurations` opção para especificar o arquivo conforme mostrado abaixo:

```
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
```

Configurando o suporte ao EMRFS S3EC V2

As SDK versões Java do S3 (1.11.837 e posteriores) oferecem suporte ao cliente de criptografia versão 2 (S3EC V2) com vários aprimoramentos de segurança. Para obter mais informações, consulte a publicação no blog do S3 [Updates to the Amazon S3 encryption client](#). Além disso, consulte a [migração do cliente de criptografia Amazon S3](#) no Guia do AWS SDK for Java desenvolvedor.

O cliente de criptografia V1 ainda está disponível no SDK para compatibilidade com versões anteriores. Por padrão, EMRFS usará o S3EC V1 para criptografar e descriptografar objetos do S3, se estiver ativado. CSE

Objetos do S3 criptografados com o S3EC V2 não podem ser descriptografados EMRFS em um EMR cluster cuja versão de lançamento seja anterior ao emr-5.31.0 (emr-5.30.1 e anterior, emr-6.1.0 e anterior).

Example Configurar EMRFS para usar o S3EC V2

Para configurar EMRFS para usar o S3EC V2, adicione a seguinte configuração:

```
{
  "Classification": "emrfs-site",
  "Properties": {
    "fs.s3.cse.encryptionV2.enabled": "true"
  }
}
```

Propriedades **emrfs-site.xml** para criptografia do lado do cliente do Amazon S3

Propriedade	Valor padrão	Descrição
<code>fs.s3.cse.enabled</code>	false	Quando definido como true , os EMRFS objetos armazenados no Amazon S3 são criptografados

Propriedade	Valor padrão	Descrição
		usando criptografia do lado do cliente.
<code>fs.s3.cse.encryptionV2.enabled</code>	false	Quando definido como <code>true</code> , EMRFS usa o cliente de criptografia S3 versão 2 para criptografar e descriptografar objetos no S3. Disponível para a EMR versão 5.31.0 e posterior.
<code>fs.s3.cse.encryptionMaterialsProvider.uri</code>	N/A	Aplica-se ao usar materiais de criptografia personalizada. O Amazon S3 URI onde o JAR with the <code>EncryptionMaterialsProvider</code> está localizado. Quando você fornece isso URI, a Amazon baixa EMR automaticamente o JAR para todos os nós do cluster.
<code>fs.s3.cse.encryptionMaterialsProvider</code>	N/A	O caminho da classe <code>EncryptionMaterialsProvider</code> usado com a criptografia no lado do cliente. Ao usar CSE -KMS, especifique <code>com.amazonaws.emr.hadoop.fs.cse.KMSEncryptionMaterialsProvider</code> .

Propriedade	Valor padrão	Descrição
<code>fs.s3.cse.materialsDescription.enabled</code>	false	Quando definido como <code>true</code> , preenche o número <code>materialsDescription</code> de objetos criptografados com o Amazon URI S3 para o objeto e o <code>JobFlowId</code> . Defina como <code>true</code> ao usar materiais de criptografia personalizados.
<code>fs.s3.cse.kms.keyId</code>	N/A	Aplica-se ao usar CSE -KMS. O valor do <code>KeyId</code> , ARN, ou alias da KMS chave usada para criptografia.
<code>fs.s3.cse.cryptoStorageMode</code>	ObjectMetadata	O modo de armazenamento do Amazon S3. Por padrão, a descrição das informações de criptografia é armazenada nos metadados do objeto. Você também pode armazenar a descrição em um arquivo de instruções. Os valores válidos são <code>ObjectMetadata</code> , <code>InstructionFile</code> e <code>InstructionFile</code> . Para obter mais informações, consulte Criptografia de dados do lado do cliente com o AWS SDK for Java e o Amazon S3 .

CloudWatch Agente da Amazon

O CloudWatch agente da Amazon na Amazon EMR é uma ferramenta que pode monitorar as EC2 instâncias da Amazon em seu EMR cluster. Você pode armazenar e visualizar as métricas coletadas com o CloudWatch agente em CloudWatch. Para obter mais informações sobre o CloudWatch agente, consulte o [Guia CloudWatch do usuário da Amazon](#).

Note

Você incorrerá em cobranças adicionais se usar outros Serviços da AWS para publicar, consultar ou visualizar métricas de CloudWatch agentes da Amazon. Consulte as páginas a seguir para obter mais informações sobre preços.

- [CloudWatch Preços da Amazon](#)
- [Preços do Amazon Managed Service para Prometheus](#)
- [Preços do Amazon Managed Grafana](#)

Essas métricas são separadas das métricas gratuitas que a Amazon EMR fornece sob o AWS/ElasticMapReduce namespace em. CloudWatch Para obter mais informações sobre as métricas que o CloudWatch agente não publica, consulte [Monitoramento de EMR métricas da Amazon com CloudWatch](#).

Com o Amazon EMR 7.0 e superior, você pode instalar uma versão personalizada do CloudWatch agente da Amazon em seu cluster para coletar métricas do seu EMR cluster. O agente coleta as métricas de todos os nós do cluster, as reúne no nó primário e publica as métricas desse nó na nuvem.

Você pode instalar o agente ao criar um novo cluster no console ou ao usar `create-cluster` API o. Para ter mais informações, consulte [Crie um EMR cluster que use o CloudWatch agente da Amazon](#). Por padrão, um EMR cluster da Amazon que executa a série versão 7.x publica [34 métricas CloudWatch em nível de sistema](#) em intervalos de 60 segundos, mas você pode configurar o agente para publicar métricas diferentes. Outra opção é publicar métricas no Amazon Managed Service for Prometheus, e você também pode escolher quais métricas publicar. Para diferentes casos de uso e configurações, você pode configurar o agente para visualizar e consultar as métricas no CloudWatch console, no Amazon Managed Grafana ou por meio APIs do for CloudWatch ou Amazon

Managed Service for Prometheus. Eles Serviços da AWS geram cobranças quando você os usa para armazenar e consultar métricas.

O CloudWatch agente na Amazon EMR pode publicar métricas do sistema, bem como JMX métricas para os seguintes serviços em seu EMR cluster da Amazon.

- Hadoop DataNode
- Hadoop NameNode
- Fio NodeManager
- Fio ResourceManager
- HBaseMaster — somente Amazon EMR 7.1+
- HBase RegionServer — Somente Amazon EMR 7.1+
- HBase ThriftServer — Somente Amazon EMR 7.1+

Para obter mais informações sobre as métricas disponíveis e como configurar o CloudWatch agente na AmazonEMR, consulte [Configurando o CloudWatch agente para a Amazon EMR](#).

A tabela a seguir lista a versão AmazonCloudWatchAgent incluída na versão mais recente da série Amazon EMR 7.x, junto com os componentes com os quais a Amazon é EMR instalada. AmazonCloudWatchAgent

Para a versão dos componentes instalados AmazonCloudWatchAgent nesta versão, consulte [Versões de componentes da versão 7.2.0](#).

AmazonCloudWatchAgent informações sobre a versão do emr-7.2.0

Gravadora EMR de lançamento da Amazon	AmazonCloudWatchAgent Versão	Componentes instalados com 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-https-server,

Gravadora EMR de lançamento da Amazon	AmazonCloudWatchAgent Versão	Componentes instalados com AmazonCloudWatchAgent
		hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

Tópicos

- [Crie um EMR cluster que use o CloudWatch agente da Amazon](#)
- [Métricas padrão para CloudWatch agentes com a Amazon EMR](#)
- [Configurando o CloudWatch agente para a Amazon EMR](#)
- [Considerações e limitações](#)
- [CloudWatch histórico de lançamentos do agente](#)

Crie um EMR cluster que use o CloudWatch agente da Amazon

Os procedimentos nesta seção descrevem as etapas para criar um cluster na Amazon EMR com o CloudWatch agente da Amazon a partir do AWS Management Console e do AWS CLI.

Tópicos

- [IAMPermissões necessárias para o CloudWatch agente](#)
- [Endpoint de CloudWatch agente necessário](#)
- [Criar um cluster do EMR](#)

IAMPermissões necessárias para o CloudWatch agente

O CloudWatch agente exige a `cloudwatch:PutMetricData` permissão AWS Identity and Access Management (IAM) no perfil de EC2 instância da Amazon para a AmazonEMR. A função EMR padrão da Amazon já tem essa permissão. Você pode criar a função padrão a partir do AWS CLI `aws emr create-default-roles`. Para obter mais informações, consulte [Função de serviço para EC2 instâncias de cluster \(perfil de EC2 instância\)](#) no Amazon EMR Management Guide.

O exemplo de IAM política a seguir inclui a `cloudwatch:PutMetricData` permissão:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "cloudwatch:PutMetricData",
      "Resource": "*"
    }
  ]
}
```

Endpoint de CloudWatch agente necessário

Para publicar métricas CloudWatch para um EMR cluster em uma sub-rede privada, crie um endpoint de CloudWatch agente e associe ao ponto em VPC que a sub-rede privada está.

Para obter mais informações sobre os CloudWatch endpoints de cada um Região da AWS, consulte os [CloudWatch endpoints e cotas da Amazon no Guia](#) de referência AWS geral.

Criar um cluster do EMR

Depois de configurar as permissões e o endpoint necessários para uso com o CloudWatch agente, use o AWS Management Console ou o AWS CLI para criar um novo cluster com o agente instalado.

Console

Para criar um cluster com o CloudWatch agente da Amazon a partir do console

1. Navegue até o EMR console da Amazon.
2. Selecione Criar cluster.
3. Em Nome e aplicativos, escolha uma EMR versão 7.0.0 ou superior da Amazon.
4. Em Pacote de aplicativos, selecione o pacote ou os aplicativos que você deseja instalar no seu cluster e inclua o CloudWatch agente nas suas seleções.
5. Prossiga com a criação do cluster para atender às necessidades do seu caso de uso.

AWS CLI

No AWS CLI, você pode adicionar o CloudWatch agente da Amazon a um cluster com o `--applications` parâmetro `paracreate-cluster`.

Para criar um cluster com o CloudWatch agente da Amazon a partir do AWS CLI

- Ao criar um cluster, use um comando semelhante ao seguinte para incluir o CloudWatch agente da Amazon. *myKey* substitua pelo nome do seu EC2 key pair.

```
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
```

Para obter mais detalhes sobre como usar a Amazon EMR com o AWS CLI, consulte a [Referência de AWS CLI Comandos](#).

Métricas padrão para CloudWatch agentes com a Amazon EMR

Quando você instala o CloudWatch agente da Amazon na AmazonEMR, a configuração padrão publica as seguintes métricas do sistema para todas as instâncias em seu cluster, a menos que você [configure o agente de forma diferente](#). Para obter as definições de cada métrica, consulte [Métricas coletadas pelo CloudWatch agente](#) no Guia CloudWatch do usuário da Amazon.

CPU

CPU métricas

- `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

Métricas de disco

- `disk_free`
- `disk_total`
- `disk_used`
- `disk_used_percent`

Memory

Métricas de memória

- `mem_active`
- `mem_available`
- `mem_available_percent`
- `mem_free`
- `mem_inactive`
- `mem_total`
- `mem_used`
- `mem_used_percent`
- `mem_buffered`
- `mem_cached`

Network IO

Métricas de E/S de rede

- `net_bytes_recv`
- `net_bytes_sent`
- `net_packets_recv`
- `net_packets_sent`

Process

Métricas de processo

- `processes_running`
- `processes_total`

Swap

Métricas de troca

- `swap_free`
- `swap_used`
- `swap_used_percent`

Por padrão, o agente publica todas as CloudWatch métricas no namespace personalizado e no esquema `CWAgentcluster.id,,, instance.id node.type service.name`. Use as etapas a seguir para visualizar essas métricas padrão que o CloudWatch agente publica para a AmazonEMR:

1. Navegue até o CloudWatch console da Amazon.
2. Escolha a guia Métricas e Todas as métricas.
3. Em Namespaces personalizados, selecione **CWAgent**. Em seguida, selecione o esquema: **cluster.id,, instance.idnode.type, service.name**.
4. Continue consultando as métricas da CloudWatch interface conforme seu caso de uso exigir.

A Amazon EMR anexa os seguintes rótulos a cada métrica: `["cluster.id", "instance.id", "node.type", "service.name"]`. Considere o seguinte em relação a esses rótulos:

- O valor do `cluster.id` rótulo é o ID do EMR cluster que produziu a métrica. Um exemplo de valor para o rótulo `cluster.id` é `j-123456789ABC`.
- O valor do `instance.id` rótulo é o ID da instância no EMR cluster que produziu a métrica. Um exemplo de valor para o rótulo `instance.id` é `i-01bcf5f140f3355777`.
- O valor do `node.type` rótulo representa o tipo de nó da instância no cluster que produziu a métrica. Esse valor pode ser `mastercore`, `outask`.
- O valor do rótulo `service.name` é o nome do serviço que produziu a métrica. O valor `service.name` padrão para as métricas padrão do sistema é `system`.

Note

A EMR versão 7.0.0 da Amazon usa um esquema de rótulo métrico ligeiramente diferente. `cluster.id` está `jobflow.id` e não `node.type` está disponível. `instance.id` e `service.name` estão disponíveis.

Configurando o CloudWatch agente para a Amazon EMR

O Amazon EMR 7.0.0 e versões posteriores incluem o CloudWatch agente Amazon. As seções a seguir descrevem como configurar o agente em seu EMR cluster, conforme descrito em [Crie um EMR cluster que use o CloudWatch agente da Amazon](#).

Tópicos

- [Configurar CloudWatch agente para Amazon EMR 7.1.0](#)
- [Configurar CloudWatch agente para Amazon EMR 7.0.0](#)

Configurar CloudWatch agente para Amazon EMR 7.1.0

A partir do Amazon EMR 7.1.0, você pode configurar o CloudWatch agente da Amazon para usar métricas adicionais do sistema, adicionar métricas de aplicativos e alterar o destino das métricas usando a EMR configuração API da Amazon. Para obter mais informações sobre como usar a EMR configuração API para configurar os aplicativos do seu cluster, consulte [Configurar aplicativos](#).

Note

7.1.0 suporta apenas o tipo de reconfiguração. `OVERWRITE` Para mais informações sobre os tipos de reconfiguração, consulte [Considerações ao reconfigurar um](#) grupo de instâncias.

Tópicos

- [Esquema de configuração](#)
- [Exemplos de configurações de métricas do sistema](#)
- [Exemplos de configurações de métricas de aplicativos](#)
- [Exemplo do Amazon Managed Service para Prometheus](#)

Esquema de configuração

emr-metricstem as seguintes classificações:

- `emr-system-metrics`— configure métricas do sistemaCPU, como disco e memória.
- `emr-hadoop-hdfs-datanode-metrics`— configurar métricas do Hadoop DataNode JMX
- `emr-hadoop-hdfs-namenode-metrics`— configurar métricas do Hadoop NameNode JMX
- `emr-hadoop-yarn-nodemanager-metrics`— configurar métricas do Yarn NodeManager JMX
- `emr-hadoop-yarn-resourcemanagemer-metrics`— configurar métricas do Yarn ResourceManager JMX
- `emr-hbase-master-metrics`— configurar JMX métricas HBase mestras
- `emr-hbase-region-server-metrics`— configurar JMX métricas do HBase Region Server
- `emr-hbase-rest-server-metrics`— configurar JMX métricas HBase REST do servidor
- `emr-hbase-thrift-server-metrics`— configurar métricas do HBase Thrift Server JMX

As tabelas a seguir descrevem as propriedades e configurações disponíveis para todas as classificações.

propriedades emr-metrics

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores	Observações
<code>metrics_destination</code>	Opcional	Determina se as métricas do cluster são publicadas na Amazon CloudWatch ou no Amazon Managed Service for Prometheus.	"CLOUDWATCH"	"CLOUDWATCH", "PROMETHEUS"	Essa propriedade não diferencia maiúsculas de minúsculas. Por exemplo, "Cloudwatch" é o mesmo que "CLOUDWATCH".

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores	Observações
<code>prometheus_endpoint</code>	Opcional	Se <code>metrics_destination</code> estiver definida como "PROMETHEUS", essa propriedade configura o CloudWatch agente para enviar métricas para o endpoint de gravação remota do Amazon Managed Service for Prometheus fornecido.	N/D	Qualquer Amazon Managed Service válido para gravação remota do Prometheus. URL O URL formato de gravação remota é <pre>https:// aps- workspaces. <region> s.com/ workspaces/ <workspace_id> / api/v1/remote_write</pre>	Esse campo é obrigatório se <code>metrics_destination</code> estiver definido como "PROMETHEUS". O provisionamento falhará se você não fornecer uma chave ou se o valor for uma string vazia.

emr-system-metrics propriedades

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores	Observações
<code>metrics_collection</code>	Opcional	Com que frequência, em	"60"	Uma string especificando o número de	Você pode substituir essa

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores	Observações
<code>_interval</code>		segundos, as métricas são coletadas e publicadas pelo CloudWatch agente.		segundos. Só aceita números inteiros.	propriedade de <code>metrics_collection_interval</code> propriedade de grupos de métricas individuais.

emr-system-metrics configurações

cpu

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores	Observações
<code>metrics</code>	Opcional	A lista de CPU métricas que o agente deve coletar.	Veja as métricas padrão para CloudWatch agentes com a Amazon EMR	Uma lista separada por vírgula de nomes de CPU métricas válidos com ou sem o <code>cpu_ prefixo</code> , como <code>usage_active</code> <code>cpu_time_idle</code> . Consulte Métricas coletadas	Especificar uma string vazia significa não publicar nenhuma CPU métrica.

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores	Observações
				pelo CloudWatch agente para ver métricas válidas.	
<code>metrics_collection_interval</code>	Opcional	Com que frequência, em segundos, o agente deve coletar e publicar CPU métricas.	O valor do <code>globalmetrics_collection_interval</code> .	Uma string especificando o número de segundos. Aceita somente números inteiros.	Esse valor substitui a <code>metrics_collection_interval</code> propriedade de <code>global</code> somente para CPU métricas.

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores	Observações
<code>drop_original_metrics</code>	Opcional	Lista de CPU métricas para as quais não publicar métricas não agregadas.	Nenhuma CPU métrica não agregada publicada.	Uma lista de CPU métricas separadas por vírgulas que também são especificadas na propriedade <code>metrics</code> . Uma string vazia significa publicar todas as CPU métricas.	O CloudWatch agente agrega todas as métricas por ID do cluster, ID da instância, tipo de nó e nome do serviço. Por padrão, o CloudWatch agente não publica as métricas por recurso para métricas com vários recursos.
<code>resources</code>	Opcional	Determina se o agente publicará métricas por núcleo.	<code>"*"</code>	<code>"*"</code> habilitar métricas por núcleo. <code>""</code> desabilitar métricas por núcleo.	O CloudWatch agente só publica métricas por núcleo para CPU métricas que não são incluídas. <code>drop_original_metrics</code>

disk

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores	Observações
<code>metrics</code>	Opcional	A lista de métricas de disco para o agente coletar.	Veja as métricas padrão para CloudWatch agentes com a Amazon EMR	Uma lista separada por vírgula de nomes de métricas de disco válidos com ou sem o <code>disk_</code> prefixo, como <code>e. disk_total used_percent</code> . Consulte Métricas coletadas pelo CloudWatch agente para ver métricas válidas.	Especificar uma string vazia significa não publicar nenhuma métrica de disco.
<code>metrics_collection_interval</code>	Opcional	Com que frequência, em segundos, o agente deve coletar e publicar métricas de disco.	O valor do <code>globalmetrics_collection_interval</code> .	Uma string especificando o número de segundos. Aceita somente números inteiros.	Esse valor substitui a <code>metrics_collection_interval</code> propriedade de <code>global</code> somente para

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores	Observações
					métricas de disco.
<code>drop_original_metrics</code>	Opcional	Lista de métricas de disco para as quais não publicar métricas não agregadas.	Nenhuma métrica de disco não agregada publicada.	Uma lista separada por vírgulas de métricas de disco que também são especificadas na propriedade <code>metrics</code> . Uma string vazia significa publicar todas as métricas do disco.	O CloudWatch agente agrega todas as métricas por ID do cluster, ID da instância, tipo de nó e nome do serviço. Por padrão, o CloudWatch agente não publica as métricas por recurso para métricas com vários recursos.

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores	Observações
<code>resources</code>	Opcional	Determina se o agente publicará per-mount-point métricas.	<code>"*"</code>	<code>"*"</code> significa todos os pontos de montagem, <code>"</code> significa nenhum ponto de montagem ou uma lista separada por vírgulas de pontos de montagem. Por exemplo, <code>"/,/emr"</code> .	O CloudWatch agente só publica per-mount-point métricas para métricas de disco que não são incluídas <code>drop_original_metrics</code> .

diskio

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores	Observações
<code>metrics</code>	Opcional	A lista de métricas de E/S do disco para o agente coletar.	Veja as métricas padrão para CloudWatch agentes com a Amazon EMR	Uma lista separada por vírgulas de nomes de métricas de E/S de disco válidos com ou sem o <code>diskio_</code> prefixo, como e. <code>diskio_re</code>	Especificar uma string vazia significa não publicar nenhuma métrica de E/S de disco.

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores	Observações
				ads writes Consulte Métricas coletadas pelo CloudWatch agente para ver métricas válidas.	
metrics_collection_interval	Opcional	Com que frequência, em segundos, o agente deve coletar e publicar métricas de E/S do disco.	O valor do globalmetrics_collection_interval.	Uma string especificando o número de segundos. Aceita somente números inteiros.	Esse valor substitui a metrics_collection_interval propriedade de global somente para métricas de E/S do disco.

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores	Observações
<code>drop_original_metrics</code>	Opcional	Lista de métricas de E/S de disco para as quais não publicar métricas não agregadas.	Nenhuma métrica de E/S de disco não agregada publicada.	Uma lista separada por vírgulas de métricas de E/S de disco que também são especificadas na propriedade <code>metrics</code> . Uma string vazia significa publicar todas as métricas de E/S do disco.	O CloudWatch agente agrega todas as métricas por ID do cluster, ID da instância, tipo de nó e nome do serviço. Por padrão, o CloudWatch agente não publica as métricas por recurso para métricas com vários recursos.

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores	Observações
<code>resources</code>	Opcional	Determina se o agente publicará métricas por dispositivo.	<code>"*"</code>	<code>"*"</code> significa todos os dispositivos de armazenamento, <code>"</code> significa nenhum dispositivo de armazenamento ou uma lista de nomes de dispositivos separados por vírgulas. Por exemplo, <code>"nvme0n1, nvme1n1"</code> .	O CloudWatch agente publica somente métricas por dispositivo para métricas de E/S de disco que não são incluídas. <code>drop_original_metrics</code>

mem

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores	Observações
<code>metrics</code>	Opcional	A lista de métricas de memória para o agente coletar.	Veja as métricas padrão para CloudWatch agentes com a Amazon EMR	Uma lista separada por vírgula de nomes de métricas de memória válidos com	Especificar uma string vazia significa não publicar nenhuma

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores	Observações
				ou sem o mem_ prefixo, como e. mem_available available_percent Consulte Métricas coletadas pelo CloudWatch agente para ver métricas válidas.	métrica de memória.
metrics_collection_interval	Opcional	Com que frequência, em segundos, o agente deve coletar e publicar métricas de memória.	O valor do globalmetrics_collection_interval.	Uma string especificando o número de segundos. Aceita somente números inteiros.	Esse valor substitui a metrics_collection_interval propriedade de global somente para métricas de memória.

net

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores	Observações
<code>metrics</code>	Opcional	A lista de métricas de rede para o agente coletar.	Veja as métricas padrão para CloudWatch agentes com a Amazon EMR	Uma lista separada por vírgula de nomes de métricas de rede válidos com ou sem o <code>net_</code> prefixo, como <code>e.net_packets_sent</code> e <code>packets_received</code> . Consulte Métricas coletadas pelo CloudWatch agente para ver métricas válidas.	Especificar uma string vazia significa não publicar nenhuma métrica de rede.
<code>metrics_collection_interval</code>	Opcional	Com que frequência, em segundos, o agente deve coletar e publicar métricas de rede.	O valor do <code>globalmetrics_collection_interval</code> .	Uma string especificando o número de segundos. Aceita somente números inteiros.	Esse valor substitui a <code>metrics_collection_interval</code> propriedade de <code>global</code> somente para

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores	Observações
					métricas de rede.
<code>drop_original_metrics</code>	Opcional	Lista de métricas de rede para as quais não publicar métricas não agregadas.	Nenhuma métrica de rede não agregada publicada.	Uma lista separada por vírgulas de métricas de rede que também são especificadas na propriedade <code>metrics</code> . Uma string vazia significa publicar todas as métricas da rede.	O CloudWatch agente agrega todas as métricas por ID do cluster, ID da instância, tipo de nó e nome do serviço. Por padrão, o CloudWatch agente não publica as métricas por recurso para métricas com vários recursos.

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores	Observações
<code>resources</code>	Opcional	Determina se o agente publicará métricas por interface.	<code>"*"</code>	<code>"*"</code> significa todas as interfaces de rede, <code>""</code> significa nenhuma interface de rede ou uma lista separada por vírgulas de nomes de interfaces. Por exemplo, <code>"eth0,eth1"</code> .	O CloudWatch agente publica somente métricas por interface para métricas de rede que não são incluídas em <code>drop_original_metrics</code> .

netstat

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores	Observações
<code>metrics</code>	Opcional	A lista de métricas de estatísticas de rede para o agente coletar.	Veja as métricas padrão para CloudWatch agentes com a Amazon EMR	Uma lista separada por vírgula de nomes de métricas de memória válidos com ou sem o <code>netstat_</code> prefixo, como e.	Especificar uma string vazia significa não publicar nenhuma métrica estatística de rede.

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores	Observações
				tcp_listen netstat_udp_socket Consulte Métricas coletadas pelo CloudWatch agente para ver métricas válidas.	
metrics_collection_interval	Opcional	Com que frequência, em segundos, o agente deve coletar e publicar métricas estatísticas de rede.	O valor do globalmetrics_collection_interval.	Uma string especificando o número de segundos. Aceita somente números inteiros.	Esse valor substitui a metrics_collection_interval propriedade de global somente para métricas estatísticas de rede.

processes

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores	Observações
metrics	Opcional	A lista de métricas do	Veja as métricas	Uma lista separada	Especificar uma

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores	Observações
		processo que o agente deve coletar.	padrão para CloudWatch agentes com a Amazon EMR	por vírgula de nomes de métricas de memória válidos com ou sem o <code>processes_</code> prefixo, como <code>processes_running</code> <code>total</code> . Consulte Métricas coletadas pelo CloudWatch agente para ver métricas válidas.	string vazia significa não publicar nenhuma métrica do processo.
<code>metrics_collection_interval</code>	Opcional	Com que frequência, em segundos, o agente deve coletar e publicar as métricas do processo do sistema.	O valor do <code>globalmetrics_collection_interval</code> .	Uma string especificando o número de segundos. Aceita somente números inteiros.	Esse valor substitui a <code>metrics_collection_interval</code> propriedade de <code>global</code> somente para métricas de processo do sistema.

swap

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores	Observações
<code>metrics</code>	Opcional	A lista de métricas de swap que o agente deve coletar.	Veja as métricas padrão para CloudWatch agentes com a Amazon EMR	Uma lista separada por vírgula de nomes de métricas de memória válidos com ou sem o <code>swap_</code> prefixo, como <code>e.swap_free</code> <code>used_percent</code> Consulte Métricas coletadas pelo CloudWatch agente para ver métricas válidas.	Especificar uma string vazia significa não publicar nenhuma métrica de troca.
<code>metrics_collection_interval</code>	Opcional	Com que frequência, em segundos, o agente deve coletar e publicar	O valor do <code>globalmetrics_collection_interval</code> .	Uma string especificando o número de segundos. Aceita somente	Esse valor substitui a <code>metrics_collection_interval</code> propriedade de global somente

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores	Observações
		métricas de troca.		números inteiros.	para métricas de swap.

emr-hadoop-hdfs-datanode-propriedades métricas

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores	
<code><custom_bean_name></code>	Opcional	N/D	O CloudWatch agente MBean desse agente deve coletar métricas, comoHadoop:service=DataNode,name=DataNodeActivity . Você pode encontrar MBean nomes de amostra e suas métricas correspondentes nos JMXAMLarquivos de exemplo da Amazon EMR versão 7.0.	Uma string contendo a lista delimitada por vírgulas de métricas associadas ao MBean. Por exemplo, BlocksCached,BlockRead .	

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores	
<code>otel.metric.export.interval</code>	Opcional	Com que frequência, em milissegundos, coletar métricas do Hadoop DataNode .	"60000"	Uma string especificando o número de milissegundos. Aceita somente números inteiros.	

emr-hadoop-hdfs-namenode-propriedades métricas

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores	
<code><custom_bean_name></code>	Opcional	N/D	O CloudWatch agente MBean desse agente deve coletar métricas, comoHadoop:service=NameNode,namespace=FSNamesystem . Você pode encontrar MBean nomes de amostra e suas métricas correspondentes nos	Uma string contendo a lista delimitada por vírgulas de métricas associadas ao MBean. Por exemplo, BlockCapacity,CapacityUsedGB .	

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores	
			JMXYAMLar quivos de exemplo da Amazon EMR versão 7.0.		
<code>otel.metric.export.interval</code>	Opcional	Com que frequência, em milissegundos, coletar métricas do Hadoop NameNode .	"60000"	Uma string especificando o número de milissegundos. Aceita somente números inteiros.	

emr-hadoop-yarn-nodemanager-propriedades métricas

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores	
<code><custom_bean_name></code>	Opcional	N/D	O CloudWatch agente MBean desse agente deve coletar métricas, como <code>Hadoop:service=NodeManager,name=NodeManagerMetrics</code> . Você pode	Uma string contendo a lista delimitada por vírgulas de métricas associadas ao MBean. Por exemplo, <code>MaxCapacity,AllocatedGB</code> .	

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores	
			encontrar MBean nomes de amostra e suas métricas correspondentes nos JMXYAML arquivos de exemplo da Amazon EMR versão 7.0.		
<code>otel.metric.export.interval</code>	Opcional	Com que frequência, em milissegundos, coletar métricas do Hadoop YARN NodeManager .	"60000"	Uma string especificando o número de milissegundos. Aceita somente números inteiros.	

emr-hadoop-yarn-resourcemanager-propriedades métricas

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores	
<code><custom_bean_name></code>	Opcional	N/D	O CloudWatch agente MBean desse agente deve coletar métricas,	Uma string contendo a lista delimitada por vírgulas de métricas associadas	

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores
			comoHadoop:service=ResourceManager,name=PartitionQueueMetrics .	ao. MBean Por exemplo, MaxCapacity,MaxCapacityVCores .
			Você pode encontrar MBean nomes de amostra e suas métricas correspondentes nos JMXYAMLarquivos de exemplo da Amazon EMR versão 7.0.	
otel.metric.export.interval	Opcional	Com que frequência, em milissegundos, coletar métricas do Hadoop YARN ResourceManager .	"60000"	Uma string especificando o número de milissegundos. Aceita somente números inteiros.

emr-hbase-master-metrics propriedades

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores
<code><custom_bean_name></code>	Opcional	N/D	O CloudWatch agente MBean desse agente deve coletar métricas, como <code>Hadoop:service=HBase,name=Master,sub=AssignmentManager</code> . Você pode encontrar MBean nomes de amostra e suas métricas correspondentes nos JMX YAML arquivos de exemplo da Amazon EMR versão 7.0.	Uma string contendo a lista delimitada por vírgulas de métricas associadas ao MBean. Por exemplo, <code>AssignFailedCount, AssignSubmittedCount</code> .
<code>otel.metric.export.interval</code>	Opcional	Com que frequência, em milissegundos, coletar métricas HBase mestras.	"60000"	Uma string especificando o número de milissegundos. Aceita somente

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores	
				números inteiros.	

emr-hbase-region-server-propriedades métricas

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores	
<code><custom_bean_name></code>	Opcional	N/D	O CloudWatch agente MBean desse agente deve coletar métricas, como <code>Hadoop:service=HBase,name=RegionServer,sub=IPC</code> . Você pode encontrar MBean nomes de amostra e suas métricas correspondentes nos JMX YAML arquivos de exemplo da Amazon EMR versão 7.0.	Uma string contendo a lista delimitada por vírgulas de métricas associadas ao MBean. Por exemplo, <code>numActiveHandler,numActivePriorityHandler</code> .	

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores
<code>otel.metric.export.interval</code>	Opcional	Com que frequência, em milissegundos, coletar métricas do Servidor de HBase Região.	"60000"	Uma string especificando o número de milissegundos. Aceita somente números inteiros.

emr-hbase-rest-server-propriedades métricas

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores
<code><custom_bean_name></code>	Opcional	N/D	O CloudWatch agente MBean desse agente deve coletar métricas, como <code>Hadoop:service=HBase,name=REST</code> . Você pode encontrar MBean nomes de amostra e suas métricas correspondentes nos JMXAMLa	Uma string contendo a lista delimitada por vírgulas de métricas associadas ao MBean. Por exemplo, <code>successfulPut, successfulScanCount</code> .

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores
			quivos de exemplo da Amazon EMR versão 7.0.	
<code>otel.metric.export.interval</code>	Opcional	Com que frequência, em milissegundos, coletar métricas do HBase Rest Server.	"60000"	Uma string especificando o número de milissegundos. Aceita somente números inteiros.

emr-hbase-thrift-server-propriedades métricas

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores
<code><custom_bean_name></code>	Opcional	N/D	O CloudWatch agente MBean desse agente deve coletar métricas, como <code>Hadoop:service=HBase,name=Thrift,sub=ThriftOn</code> e . Você pode encontrar	Uma string contendo a lista delimitada por vírgulas de métricas associadas ao MBean. Por exemplo, <code>BatchGet_max,BatchGet_mean</code> .

Propriedade	Obrigatório	Descrição	Valor padrão	Possíveis valores
			MBean nomes de amostra e suas métricas correspon dentes nos JMX YAML arquivos de exemplo da Amazon EMR versão 7.0.	
otel.metric.export.interval	Opcional	Com que frequência, em milissegundos, coletar as métricas do servidor HBase Thrift.	"60000"	Uma string especificando o número de milissegundos. Aceita somente números inteiros.

Exemplos de configurações de métricas do sistema

O exemplo a seguir demonstra como configurar o CloudWatch agente para parar de exportar todas as métricas do sistema.

```
[
  {
    "Classification": "emr-metrics",
    "Properties": {},
    "Configurations": [
      {
        "Classification": "emr-system-metrics",
        "Properties": {},
        "Configurations": []
      }
    ]
  }
]
```

```

    }
  ]
}
]
```

O exemplo a seguir configura o CloudWatch agente para exportar as métricas padrão do sistema. Fazer isso é uma maneira rápida de fazer com que o agente volte a exportar apenas as métricas padrão do sistema se você já tiver reconfigurado as métricas do sistema pelo menos uma vez. Essa redefinição também remove todas as métricas do aplicativo que foram reconfiguradas anteriormente.

```

[
  {
    "Classification": "emr-metrics",
    "Properties": {},
    "Configurations": []
  }
]
```

O exemplo a seguir configura o cluster para exportar o cpumem, e as disk métricas.

```

[
  {
    "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": ""
      }
    }
  ]
}
]
```

O exemplo de configuração anterior tem as seguintes propriedades:

- A cada 30 segundos, o agente coleta a `cpu_guest` métrica para todos CPUs. Você pode encontrar a métrica agregada no CloudWatch CWAgent > `cluster.id`, `instance.id`, `node.type`, `service.name namespace`.
- A cada 30 segundos, o agente coleta a `cpu_idle` métrica para todos CPUs. Você pode encontrar a métrica agregada no CloudWatch CWAgent > `cluster.id`, `instance.id`, `node.type`, `service.name namespace`. O agente também coleta as métricas por CPU. Você pode encontrá-los no mesmo namespace. O agente coleta essa métrica porque a `drop_original_metrics` propriedade não contém `cpu_idle`, então o agente não ignora a métrica.
- A cada 20 segundos, o agente coleta a `mem_active` métrica. Você pode encontrar a métrica agregada no CloudWatch CWAgent > `cluster.id`, `instance.id`, `node.type`, `service.name namespace`.
- A cada 20 segundos, o agente coleta as `disk_used_percent` métricas das / montagens /mnt em disco. Você pode encontrar as métricas agregadas no CloudWatch CWAgent > `cluster.id`, `instance.id`, `node.type`, `service.name namespace`. O agente também coleta as métricas por montagem. Você pode encontrá-los no mesmo namespace. O agente coleta essa métrica porque a `drop_original_metrics` propriedade não contém `disk_used_percent`, então o agente não ignora a métrica.

Exemplos de configurações de métricas de aplicativos

O exemplo a seguir configura o CloudWatch agente para parar de exportar métricas para o serviço Hadoop Namenode.

```
[
  {
    "Classification": "emr-metrics",
    "Properties": {},
    "Configurations": [
      {
        "Classification": "emr-hadoop-hdfs-namenode-metrics",
        "Properties": {},
        "Configurations": []
      }
    ]
  }
]
```

O exemplo a seguir configura um cluster para exportar métricas do aplicativo 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": []
  }
]
}
]

```

O exemplo anterior tem as seguintes propriedades:

- A cada 20 segundos, o agente coleta as CapacityUsedGB métricas BlockCapacity e das instâncias que executam o serviço Hadoop Namenode.
- A cada 30 segundos, o agente coleta MemNonHeapUsedM métricas das instâncias que executam o serviço Hadoop Datanode.
- A cada 30 segundos, o agente coleta as NodeUpdateNumOps métricas AllocateNumOps e das instâncias que executam o YARN ResourceManager Hadoop.

Exemplo do Amazon Managed Service para Prometheus

O exemplo a seguir demonstra como configurar o CloudWatch agente para exportar métricas para o Amazon Managed Service for Prometheus.

Se você está atualmente exportando métricas para o Amazon Managed Service for Prometheus e quiser reconfigurar as métricas para o cluster e continuar exportando métricas para o Amazon Managed Service for Prometheus, você deve incluir as propriedades `metrics_destination` e `prometheus_endpoint`

```

[
  {
    "Classification": "emr-metrics",
    "Properties": {
      "metrics_destination": "prometheus",
      "prometheus_endpoint": "http://amp-workspace/api/v1/remote_write"
    },
    "Configurations": []
  }
]

```

Para usar o CloudWatch agente para exportar métricas CloudWatch, use o exemplo a seguir.

```
[
  {
    "Classification": "emr-metrics",
    "Properties": {
      "metrics_destination": "cloudwatch"
    },
    "Configurations": []
  }
]
```

Note

O CloudWatch agente tem um exportador Prometheus que renomeia determinados atributos. Para os rótulos de métricas padrão, o Amazon Managed Service for Prometheus usa caracteres de sublinhado no lugar dos períodos que a Amazon usa. CloudWatch Se você usar o Amazon Managed Grafana para visualizar as métricas padrão no Amazon Managed Service for Prometheus, os rótulos aparecerão como,, e. `cluster_id` `instance_id` `node_type` `service_name`

Configurar CloudWatch agente para Amazon EMR 7.0.0

Você pode configurar o CloudWatch agente da Amazon para usar métricas de sistema adicionais além daquelas [fornecidas pela configuração padrão do CloudWatch agente](#). A configuração da versão 7.0.0 requer o uso de ações de bootstrap, das quais fornecemos exemplos nas seções a seguir. Em uma próxima versão, a Amazon EMR fornecerá opções adicionais de configuração por meio da Amazon EMR API.

Tópicos

- [Configure métricas adicionais do sistema com o Amazon EMR 7.0.0](#)
- [Configure métricas de aplicativos com o Amazon EMR 7.0.0](#)
- [Configure o Amazon Managed Service para Prometheus como armazenamento em nuvem para métricas com o Amazon 7.0.0 EMR](#)

Configure métricas adicionais do sistema com o Amazon EMR 7.0.0

Use as etapas a seguir para configurar o agente para usar um conjunto diferente de métricas do sistema na Amazon EMR 7.0.0:

1. Crie ou escolha um bucket na sua conta do Amazon S3 onde você deseja armazenar os arquivos de configuração que especificam as métricas do CloudWatch agente.
2. Crie o arquivo de configuração `emr-amazon-cloudwatch-agent.json` com suas métricas preferenciais especificadas. Para fazer isso, use um dos métodos explicados em [Criar o arquivo de configuração do CloudWatch agente](#). Para obter mais informações sobre a estrutura do arquivo de configuração do CloudWatch agente, consulte [Criar ou editar manualmente o arquivo de configuração do CloudWatch agente](#) no Guia CloudWatch do usuário da Amazon.
3. Em seguida, navegue até o `aws-emr-utilities` repositório GitHub e baixe os seguintes scripts de métricas do sistema:
 - [install_system_metrics_launcher.sh](#): um script que baixa e executa `install_system_metrics.sh` em segundo plano para que o nó possa concluir o bootstrapping.
 - [install_system_metrics.sh](#)— Um script que espera que a instância em que é executado conclua a inicialização e, em seguida, baixa e aplica a configuração no arquivo. JSON
4. Abra cada arquivo SH e substitua `my-s3-bucket` pelo nome do seu bucket da Etapa 1.
5. Faça upload de um JSON e dois arquivos SH em seu bucket do S3.
6. Agora, você pode navegar até o EMR console da Amazon e criar um novo cluster com o CloudWatch agente. Em Ativado, EMREC2na navegação à esquerda, selecione Clusters e, em seguida, Criar cluster.
7. Na seção Nome e aplicativos, escolha uma EMR versão 7.0.0 ou superior da Amazon.
8. Em Pacote de aplicativos, selecione o pacote ou aplicativos que você deseja instalar em seu cluster e inclua o Amazon CloudWatch Agent em suas seleções.
9. Na seção Ações de bootstrap, selecione Adicionar.
 - Em Nome, insira `install_system_metrics_launcher.sh`.
 - Na Localização do script, insira `s3://my-s3-bucket/install_system_metrics_launcher.sh`. Substitua `my-s3-bucket` pelo caminho para o bucket do S3.
 - Deixe o bloco Argumentos vazio.

10. Selecione Adicionar ação de bootstrap.
11. Continue criando o cluster para atender às necessidades da sua workload.

Quando seu cluster é iniciado, o CloudWatch agente publica as métricas do sistema que você especificou no arquivo de configuração para CloudWatch.

Configure métricas de aplicativos com o Amazon EMR 7.0.0

Você pode configurar o CloudWatch agente da Amazon para publicar métricas de aplicativos para HDFS e YARN além das métricas do sistema. Use as etapas a seguir para configurar o agente para publicar as métricas de aplicações.

1. Crie ou escolha um bucket na sua conta do Amazon S3 onde você deseja armazenar os arquivos de configuração que especificam as métricas do CloudWatch agente.
2. Em seguida, navegue até o `aws-emr-utilities` repositório GitHub e baixe os seguintes scripts:
 - [`install_app_metrics_launcher.sh`](#): um script que baixa e executa `install_app_metrics.sh` em segundo plano para que o nó possa concluir o bootstrapping.
 - [`install_app_metrics.sh`](#)— Um script que espera que a instância em que é executado conclua a inicialização e, em seguida, baixa e aplica a configuração nos YAML arquivos que você baixará em uma próxima etapa.
3. Abra cada arquivo e substitua `my-s3-bucket` pelo nome do seu bucket da Etapa 1.
4. Em seguida, baixe os seguintes arquivos de YAML mapeamento. Para obter informações sobre como esses YAML arquivos são estruturados, consulte [javaagent](#) no OpenTelemetry Instrumentation for Java GitHub repositório.
 - [`datanode-metrics.yaml`](#): a configuração das métricas de Hadoop DataNode.
 - [`namenode-metrics.yaml`](#): a configuração das métricas de Hadoop NameNode.
 - [`nodemanager-metrics.yaml`](#): a configuração das métricas de Yarn NodeManager.
 - [`resourcemanager-metrics.yaml`](#): a configuração das métricas de Yarn ResourceManager.
5. Faça upload dos dois arquivos SH e YAML dos quatro arquivos em seu bucket do S3.
6. Agora, você pode navegar até o EMR console da Amazon e criar um novo cluster com o CloudWatch agente. Em Ativado, EMREC2 na navegação à esquerda, selecione Clusters e, em seguida, Criar cluster.

7. Na seção Nome e aplicativos, escolha uma EMR versão 7.0.0 ou superior da Amazon.
8. Em Pacote de aplicativos, selecione o pacote ou grupo personalizado de aplicativos que você deseja instalar no seu cluster e inclua o CloudWatch agente em suas seleções.
9. Na seção Ações de bootstrap, selecione Adicionar.
 - Em Nome, insira **install_app_metrics_launcher.sh**.
 - Na Localização do script, insira **s3://my-s3-bucket/install_app_metrics_launcher.sh**. Substitua *my-s3-bucket* pelo caminho para o bucket do S3.
 - Deixe o bloco Argumentos vazio.
10. Selecione Adicionar ação de bootstrap.
11. Continue criando o cluster para atender às necessidades da sua workload.

Quando seu cluster é iniciado, o CloudWatch agente publica as métricas do aplicativo que você especificou junto com as métricas do sistema para CloudWatch.

Configure o Amazon Managed Service para Prometheus como armazenamento em nuvem para métricas com o Amazon 7.0.0 EMR

Você pode configurar o CloudWatch agente da Amazon para publicar métricas no Amazon Managed Service for Prometheus em vez de CloudWatch

Note

Você pode publicar métricas do CloudWatch agente da Amazon no Amazon Managed Service for Prometheus ou na CloudWatch Amazon, mas não pode publicar as métricas nos dois serviços do mesmo cluster.

Para configurar o agente para publicar métricas no Amazon Managed Service for Prometheus, você deve adicionar `aps:RemoteWrite` AWS Identity and Access Management a permissão IAM () ao perfil de instância da Amazon para a EC2 Amazon. EMR O seguinte exemplo de política contém a permissão necessária:

```
{  
  "Version": "2012-10-17",
```

```
"Statement": [  
  {  
    "Effect": "Allow",  
    "Action": "aps:RemoteWrite",  
    "Resource": "*"  
  }  
]  
}
```

Use o CloudWatch agente em um EMR cluster para publicar métricas no Amazon Managed Service for Prometheus

Depois que a política de serviço tiver as permissões corretas, use as etapas a seguir para iniciar um cluster que usa o CloudWatch agente para publicar métricas no Amazon Managed Service for Prometheus.

1. Use o AWS Management Console ou AWS CLI para criar um espaço de trabalho do Amazon Managed Service para Prometheus. Para obter mais informações, consulte [Criação de um espaço de trabalho](#) no Guia do usuário do Amazon Managed Service for Prometheus.
2. Crie ou escolha um bucket na sua conta do Amazon S3 onde deseja armazenar os arquivos de execução que especificam o Amazon Managed Service for Prometheus como armazenamento em nuvem.
3. Em seguida, navegue até o `aws-emr-utilities` repositório GitHub e baixe os seguintes scripts:
 - [add_prometheus_endpoint_launcher.sh](#): um script que baixa e executa `add_prometheus_endpoint.sh` em segundo plano para que o nó possa concluir o bootstrapping.
 - [add_prometheus_endpoint.sh](#)— Um script que espera que a instância em que é executado conclua a inicialização e, em seguida, configura o CloudWatch agente para publicar no endpoint do Amazon Managed Service for Prometheus, que você fornece como argumento ao iniciar seu cluster.
4. Abra cada arquivo e substitua `my-s3-bucket` pelo nome do seu bucket da Etapa 2.
5. Use o AWS CLI para criar um EMR cluster com a ação `add_prometheus_endpoint_launcher.sh bootstrap`. No comando a seguir, substitua `my-s3-bucket` pelo bucket que contém a ação de bootstrap e substitua `managedpro-remote-write-workspace-url` pelo endpoint de gravação remota do seu espaço de trabalho do Amazon Managed Service for Prometheus. Certifique-se de especificar uma etiqueta de EMR lançamento da Amazon igual `emr-7.0.0` ou superior.

```
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'
```

Quando seu cluster é iniciado, o CloudWatch agente publica as métricas coletadas no Amazon Managed Service for Prometheus.

Como usar o Amazon Managed Service for Prometheus como fonte de dados do Amazon Managed Grafana

Depois que a Amazon EMR publicar as métricas do cluster no Amazon Managed Service for Prometheus, você poderá usar as seguintes etapas para visualizar as métricas com o Amazon Managed Grafana:

1. Use o AWS Management Console para criar um espaço de trabalho e um usuário do Amazon Managed Grafana com as permissões apropriadas. Para obter mais informações, consulte [Criação de um espaço de trabalho](#) no Guia do usuário do Amazon Managed Grafana.
2. Adicione seu espaço de trabalho do Amazon Managed Service for Prometheus como fonte de dados no Amazon Managed Grafana. Para obter mais informações, consulte [Use AWS data source configuration to add Amazon Managed Service for Prometheus as a data source](#) no Guia do usuário do Amazon Managed Grafana.

Note

O CloudWatch agente tem um exportador Prometheus que renomeia determinados atributos. Para os rótulos de métricas padrão, o Amazon Managed Service for Prometheus usa caracteres de sublinhado no lugar dos períodos que a Amazon usa. CloudWatch Portanto, se você usar o Amazon Managed Grafana para visualizar as métricas padrão no Amazon Managed Service for Prometheus, os rótulos aparecerão como `jobflow_id`, `instance_id` e `service_name`.

Além disso, qualquer métrica de aplicativo que o CloudWatch agente publica no Amazon Managed Service for Prometheus usa o rótulo em vez de. `job service_name` No entanto, as métricas do sistema continuam usando o rótulo `service_name`.

Considerações e limitações

- O CloudWatch agente nativo da Amazon está disponível para clusters que você cria com as EMR versões 7.0.0 e superiores da Amazon.
- A configuração do CloudWatch agente no Amazon EMR 7.0.0 requer o uso de ações de [bootstrap](#). Em uma próxima versão, a Amazon EMR fornecerá opções adicionais de configuração por meio da Amazon EMR API.
- Você não pode instalar o EMR CloudWatch agente da Amazon se já tiver implantado o CloudWatch agente por outro método, como AWS Systems Manager Agente (SSMAgente) na região em que você criou seu cluster. Isso resultará no encerramento do cluster com erros.

CloudWatch histórico de lançamentos do agente

A tabela a seguir lista a versão AmazonCloudWatchAgent incluída em cada versão de lançamento da AmazonEMR, junto com os componentes instalados com o aplicativo. Para obter as versões dos componentes em cada versão, consulte a seção Versão de componentes da sua versão em [Versões de lançamento do Amazon EMR 7.x](#), [Versões de lançamento do Amazon EMR 6.x](#) ou [Versões de lançamento do Amazon EMR 5.x](#).

AmazonCloudWatchAgent informações sobre a versão

Gravadora Amazon EMR Release	AmazonCloudWatchAgent Versão	Componentes instalados com 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,

Gravadora Amazon EMR Release	AmazonCloudWatchAgent Versão	Componentes instalados com AmazonCloudWatchAgent
		hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, 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-https-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

O Delta Lake é um framework de camada de armazenamento para arquiteturas de lakehouse comumente desenvolvidas no Amazon S3. Com as EMR versões 6.9.0 e superiores da Amazon, você pode usar o [Apache Spark 3.x](#) em EMR clusters da Amazon com tabelas Delta Lake. Para obter mais informações sobre lakehouses com Delta Lake, consulte <https://delta.io/>.

A tabela a seguir lista a versão da Delta incluída na versão mais recente da série Amazon EMR 7.x, junto com os componentes que a Amazon EMR instala com a Delta.

Para a versão dos componentes instalados com o Delta nesta versão, consulte [Versões de componentes da versão 7.2.0](#).

Informações sobre a versão Delta do emr-7.2.0

Gravadora EMR de lançamento da Amazon	Versão Delta	Componentes instalados com o Delta
emr-7.2.0	Delta 3.1.0	Not available.

A tabela a seguir lista a versão da Delta incluída na versão mais recente da série EMR 6.x da Amazon, junto com os componentes que a Amazon EMR instala com a Delta.

Para obter a versão dos componentes instalados com o Delta nessa versão, consulte [Release 6.15.0 Component Versions](#).

Informações de versão do Delta para o emr-6.15.0

Gravadora EMR de lançamento da Amazon	Versão Delta	Componentes instalados com o Delta
emr-6.15.0	Delta 2.4.0	Not available.

Introdução ao Delta Lake

O Delta Lake é um projeto de código aberto que ajuda a implementar arquiteturas modernas de data lake comumente desenvolvidas no Amazon S3. O Delta Lake oferece os seguintes recursos:

- Transações atômicas, consistentes, isoladas e duráveis (ACID) no Spark. Os leitores têm uma visualização consistente da tabela durante um trabalho do Spark.
- Tratamento escalável de metadados com processamento distribuído pelo Spark.
- Combina casos de uso de transmissão e em lote com a mesma tabela Delta.
- Aplicação automática de esquema para evitar registros incorretos durante a ingestão de dados.
- Viagem no tempo com versionamento de dados.
- Suporta operações de mesclagem, atualização e exclusão para casos de uso complexos, como captura de dados de alteração (CDC), atualizações de streaming e muito mais.

Usar um cluster com o Delta Lake instalado

Tópicos

- [Usar um cluster do Delta Lake com o Flink](#)
- [Usar um cluster Delta Lake com o Trino](#)
- [Usar um cluster do Delta Lake com o Spark](#)
- [Usar um cluster do Delta Lake com Spark e AWS Glue](#)

Usar um cluster do Delta Lake com o Flink

Com a EMR versão 6.11 e superior da Amazon, você pode usar o Delta Lake com seu cluster Flink. Os exemplos a seguir usam o AWS CLI para trabalhar com o Delta Lake em um cluster Amazon EMR Flink.

Note

A Amazon EMR oferece suporte ao Flink DataStream API quando você usa o Delta Lake com um cluster do Flink.

Criar um cluster Delta Lake

1. Crie um arquivo `delta_configurations.json`, com o seguinte conteúdo:

```
[{"Classification":"delta-defaults",
```

```
"Properties":{"delta.enabled":"true"}}}]
```

2. Crie um cluster com a configuração a seguir. Substitua a `example Amazon S3 bucket path` e o `subnet ID` por seus próprios.

```
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
```

Inicializar uma sessão do yarn do Flink

Para inicializar uma sessão do yarn do Flink, execute o seguinte comando:

```
flink-yarn-session -d
```

Compilar um trabalho no Flink com Delta Lake

Os exemplos a seguir mostram como usar o `sbt` ou o `Maven` para criar seu trabalho no Flink com o `Delta Lake`.

`sbt`

O [sbt](#) é uma ferramenta de build para Scala que pode ser usada com pouca ou nenhuma configuração quando você tem pequenos projetos.

```
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

O [Maven](#) é uma ferramenta de automação de build de código aberto da Apache Software Foundation. Com o Maven, você pode criar, publicar e implantar um trabalho do Flink com o Delta Lake na 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>

```

Grave em uma tabela Delta com o Flink Datastream API

Use o exemplo a seguir para criar um DeltaSink para gravar na tabela com um `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;
}

```

Leia de uma tabela Delta com o Flink Datastream API

Use o exemplo a seguir para criar um limite DeltaSource para leitura da tabela com um `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");
}
```

Criação de coletores com suporte a vários clusters para Delta Lake independente

Use o exemplo a seguir para criar uma DeltaSink tabela para gravar com [suporte a deltaTablePath e vários 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;
}
```

Executar o trabalho do Flink

Use o seguinte comando para executar o trabalho:

```
flink run FlinkJob.jar
```

Usar um cluster Delta Lake com o Trino

Com as EMR versões 6.9.0 e superiores da Amazon, você pode usar o Delta Lake com seu cluster Trino.

Neste tutorial, usaremos o AWS CLI para trabalhar com o Delta Lake no cluster Amazon EMR Trino.

Criar um cluster Delta Lake

1. Crie um arquivo, `delta_configurations.json`, e defina valores para o catálogo escolhido. Por exemplo, para usar o Hive Metastore como catálogo, o arquivo deve ter o seguinte conteúdo:

```
[{"Classification":"delta-defaults",
  "Properties":{"delta.enabled":"true"}},
 {"Classification":"trino-connector-delta",
  "Properties":{"hive.metastore.uri":"thrift://localhost:9083"}}]
```

Se você quiser usar o AWS Glue Catalog como sua loja, seu arquivo deve ter o seguinte conteúdo:

```
[{"Classification":"delta-defaults",
  "Properties":{"delta.enabled":"true"}},
 {"Classification":"trino-connector-delta",
  "Properties":{"hive.metastore":"glue"}}]
```

2. Crie um cluster com a configuração a seguir, substituindo **example Amazon S3 bucket path** e **subnet ID** por suas informações.

```
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
```

Inicializar a sessão do Trino para Delta Lake

Inicializar uma sessão do Trino, execute o comando a seguir

```
trino-cli --catalog delta
```

Gravar em uma tabela do Delta Lake

Crie e grave na sua tabela com os seguintes SQL comandos:

```
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');
```

Ler uma tabela do Delta Lake

Leia da sua tabela com o seguinte SQL comando:

```
SELECT * from default.delta_table;
```

Usar um cluster do Delta Lake com o Spark

A partir da EMR versão 6.9.0 da Amazon, você pode usar o Delta Lake com seu cluster Spark sem a necessidade de ações de bootstrap. Para as EMR versões 6.8.0 e inferiores da Amazon, você pode usar ações de bootstrap para pré-instalar as dependências necessárias.

Os exemplos a seguir usam o AWS CLI para trabalhar com o Delta Lake em um cluster do Amazon EMR Spark.

Para usar o Delta Lake na Amazon EMR com o AWS Command Line Interface, primeiro crie um cluster. Para obter informações sobre como especificar a classificação Delta Lake com AWS Command Line Interface, consulte [Forneça uma configuração usando o AWS Command Line Interface ao criar um cluster](#) ou [Forneça uma configuração com o Java SDK ao criar um cluster](#).

1. Crie um arquivo `configurations.json`, com o seguinte conteúdo:

```
[{"Classification":"delta-defaults", "Properties":{"delta.enabled":"true"} }]
```

2. Crie um cluster com a configuração a seguir, substituindo o **bucket path** do Amazon S3 de exemplo e a **subnet ID** por suas informações.

```
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
```

Como alternativa, você pode criar um EMR cluster Amazon e um aplicativo Spark com os seguintes arquivos como JAR dependências em um trabalho do 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

Se você usa as EMR versões 6.9.0 ou superiores da Amazon, use `/usr/share/aws/delta/lib/delta-spark.jar` em vez de `/usr/share/aws/delta/lib/delta-core.jar`

Para obter mais informações, consulte [Submitting Applications](#).

Para incluir uma dependência de jar no trabalho do Spark, é possível adicionar as seguintes propriedades de configuração à aplicação 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"
```

Para obter mais informações sobre as dependências de trabalhos do Spark, consulte [Dependency Management](#).

Se você usa as EMR versões 6.9.0 ou superiores da Amazon, adicione a `/usr/share/aws/delta/lib/delta-spark.jar` configuração em vez disso.

```
--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"
```

Inicializar uma sessão do Spark para Delta Lake

Os exemplos a seguir mostram como iniciar o shell interativo do Spark, usar o envio do Spark ou usar o Amazon EMR Notebooks para trabalhar com o Delta Lake na Amazon. EMR

spark-shell

1. Conecte-se ao nó primário usando SSH o. Para obter mais informações, consulte [Conecte-se ao nó primário usando SSH](#) o Amazon EMR Management Guide.
2. Digite o seguinte comando para iniciar o shell do Spark. Para usar a PySpark concha, `spark-shell` substitua `porpyspark`.

```
spark-shell \
  --conf "spark.sql.extensions=io.delta.sql.DeltaSparkSessionExtension" \
  --conf
  "spark.sql.catalog.spark_catalog=org.apache.spark.sql.delta.catalog.DeltaCatalog"
```

Se você executa as EMR versões 6.15.0 ou superiores da Amazon, você também deve usar as seguintes configurações para usar um controle de acesso refinado baseado em Lake Formation com 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. Conecte-se ao nó primário usando SSH o. Para obter mais informações, consulte [Conecte-se ao nó primário usando SSH](#) o Amazon EMR Management Guide.
2. Insira o comando a seguir para iniciar a sessão do Spark no 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"
```

Se você executa as EMR versões 6.15.0 ou superiores da Amazon, você também deve usar as seguintes configurações para usar um controle de acesso refinado baseado em Lake Formation com 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

Para inicializar uma sessão do Spark usando os notebooks do Amazon EMR Studio, configure sua sessão do Spark usando o comando `%%configure` mágico no seu EMR notebook Amazon, como no exemplo a seguir. Para obter mais informações, consulte [Use EMR Notebooks magics](#) no Amazon EMR Management Guide.

```
%%configure -f
{
  "conf": {
    "spark.sql.extensions": "io.delta.sql.DeltaSparkSessionExtension",
    "spark.sql.catalog.spark_catalog":
"org.apache.spark.sql.delta.catalog.DeltaCatalog"
  }
}
```

Se você executa as EMR versões 6.15.0 ou superiores da Amazon, você também deve usar as seguintes configurações para usar um controle de acesso refinado baseado em Lake Formation com 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"
  }
}
```

Gravar em uma tabela do Delta Lake

O exemplo a seguir mostra como criar um DataFrame e gravá-lo como um conjunto de dados do Delta Lake. O exemplo mostra como trabalhar com conjuntos de dados com o shell do Spark enquanto estiver conectado ao nó primário usando SSH como usuário padrão do hadoop.

Note

Para colar exemplos de código no shell do Spark, digite `:paste` no prompt, cole o exemplo e pressione CTRL + D.

PySpark

O Spark inclui um shell baseado em Python, `pyspark`, que você pode usar para gerar protótipos de programas Spark gravados em Python. Assim como com `spark-shell`, invoque `pyspark` no nó primário.

```
## 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");

```

Ler uma tabela do Delta Lake

PySpark

```

ddf = spark.table("delta_table")
ddf.show()

```

Scala

```

val ddf = spark.table("delta_table")
ddf.show()

```

SQL

```
SELECT * FROM delta_table;
```

Usar um cluster do Delta Lake com Spark e AWS Glue

Para usar o AWS Glue Catalog como Metastore para tabelas Delta Lake, crie um cluster com as etapas a seguir. Para obter informações sobre como especificar a classificação Delta Lake usando AWS Command Line Interface, consulte [Forneça uma configuração usando o AWS Command Line Interface ao criar um cluster](#) ou [Forneça uma configuração usando o Java SDK ao criar um cluster](#).

Criar um cluster Delta Lake

1. Crie um arquivo `configurations.json`, com o seguinte conteúdo:

```
[{"Classification":"delta-defaults",
  "Properties":{"delta.enabled":"true"}},
{"Classification":"spark-hive-site",
  "Properties":
  {"hive.metastore.client.factory.class":"com.amazonaws.glue.catalog.metastore.AWSGlueDataCat
```

2. Crie um cluster com a configuração a seguir, substituindo **example Amazon S3 bucket path** e **subnet ID** por suas informações.

```
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
```

Considerações e limitações

- O Delta Lake é compatível com as EMR versões 6.9.0 e superiores da Amazon. Você pode usar o [Apache Spark 3.x](#) em EMR clusters da Amazon com tabelas Delta.
- Recomendamos que você use o URI esquema s3 para caminhos de localização do S3 em vez do s3a para obter melhor desempenho, segurança e confiabilidade. Para obter mais informações, consulte [Working with storage and file systems](#).
- Com o Amazon EMR 7.0, Delta Universal Format (UniForm) as declarações Convert-to-Iceberg não são suportadas.
- Com o Amazon EMR 6.9 e 6.10, quando você armazena dados da tabela Delta Lake no Amazon S3, os dados da coluna NULL se tornam após a operação de renomeação da coluna. Esse problema foi resolvido com o Amazon EMR 6.11. Para obter mais informações sobre a operação experimental de renomeação de coluna, consulte [Column rename operation](#) no guia do usuário do Delta Lake.
- Se você criar um banco de dados no AWS Glue Data Catalog fora do Apache Spark, o banco de dados poderá ter um campo vazioLOCATION. Como o Spark não permite a criação de bancos de dados com uma propriedade de localização vazia, você receberá o seguinte erro se usar o Spark na Amazon EMR para criar uma tabela Delta em um banco de dados Glue e o banco de dados tiver uma propriedade vaziaLOCATION:

```
IllegalArgumentException: Can not create a Path from an empty string
```

Para resolver esse problema, crie o banco de dados no Catálogo de Dados com um caminho válido e não vazio para o campo LOCATION. Para ver as etapas para implementar a solução, consulte [Exceção para argumento inválido ao criar uma tabela](#) no Guia do usuário do Amazon Athena.

Histórico de versões do Delta

A tabela a seguir lista a versão do Delta incluída em cada versão de lançamento da AmazonEMR, junto com os componentes instalados com o aplicativo. Para obter as versões dos componentes em cada versão, consulte a seção Versão de componentes da sua versão em [Versões de lançamento do Amazon EMR 7.x](#), [Versões de lançamento do Amazon EMR 6.x](#) ou [Versões de lançamento do Amazon EMR 5.x](#).

Informações da versão Delta

Gravadora Amazon EMR Release	Versão Delta	Componentes instalados com o 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

O [Apache Flink](#) é um mecanismo de fluxo de dados de streaming que você pode usar para executar o processamento de streaming em tempo real em fontes de dados de alto throughput. O Flink suporta semântica de tempo de evento para out-of-order eventos, semântica de exatamente uma vez, controle de contrapressão e APIs otimizado para escrever aplicativos de streaming e em lote.

Além disso, o Flink tem conectores para fontes de dados de terceiros, como o seguinte:

- [Amazon Kinesis Data Streams](#)
- [Apache Kafka](#)
- [Elasticsearch Connector do Flink](#)
- [Transmissão no Twitter API](#)
- [Cassandra](#)

A Amazon EMR oferece suporte ao Flink como um YARN aplicativo para que você possa gerenciar recursos junto com outros aplicativos em um cluster. O Flink-on- YARN permite que você envie trabalhos temporários do Flink ou crie um cluster de longa execução que aceite vários trabalhos e aloque recursos de acordo com a reserva geral. YARN

O Flink está incluído nas versões 5.1.0 e posteriores da AmazonEMR.

Note

Support para a `FlinkKinesisConsumer` classe foi adicionado na EMR versão 5.2.1 da Amazon.

A tabela a seguir lista a versão do Flink incluída na versão mais recente da série Amazon EMR 7.x, junto com os componentes que a Amazon EMR instala com o Flink.

Para a versão dos componentes instalados com o Flink nesta versão, consulte [Versões de componentes da versão 7.2.0](#).

Informações sobre a versão do Flink para o emr-7.2.0

Gravadora EMR de lançamento da Amazon	Versão do Flink	Componentes instalados com o 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

A tabela a seguir lista a versão do Flink incluída na versão mais recente da série EMR 6.x da Amazon, junto com os componentes que a Amazon EMR instala com o Flink.

Para obter a versão dos componentes instalados com o Flink nessa versão, consulte [Release 6.15.0 Component Versions](#).

Informações de versão do Flink para o emr-6.15.0

Gravadora EMR de lançamento da Amazon	Versão do Flink	Componentes instalados com o 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,

Gravadora EMR de lançamento da Amazon	Versão do Flink	Componentes instalados com o Flink
		flink-jobmanager-config, hudi, delta-standalone-connectors

A tabela a seguir lista a versão do Flink incluída na versão mais recente da série Amazon EMR 5.x, junto com os componentes que a Amazon EMR instala com o Flink.

Para a versão dos componentes instalados com o Flink nesta versão, consulte [Versões de componentes da versão 5.36.2](#).

Informações sobre a versão do Flink para o emr-5.36.2

Gravadora EMR de lançamento da Amazon	Versão do Flink	Componentes instalados com o 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

Tópicos

- [Criar um cluster com o Flink](#)
- [Configurando o Flink na Amazon EMR](#)
- [Usar trabalhos do Flink no Amazon EMR](#)
- [Usar o shell Scala](#)
- [Localizar a interface Web do Flink](#)
- [Escalador automático do Flink](#)

- [Otimização dos tempos de reinício dos trabalhos para operações de recuperação e ajuste de escala de tarefas](#)
- [Usar trabalhos do Flink pelo Zeppelin no Amazon EMR](#)
- [Histórico de versões do Flink](#)

Criar um cluster com o Flink

Você pode iniciar um cluster com o AWS Management Console AWS CLI, ou um AWS SDK.

Iniciar um cluster com o Flink instalado no console

1. Abra o EMR console da Amazon em <https://console.aws.amazon.com/emr>.
2. Escolha Create cluster (Criar cluster), Go to advanced options (Ir para opções avançadas).
3. Para Configuração de software, escolha EMRRelease emr-5.1.0 ou posterior.
4. Selecione Flink como um aplicativo, juntamente com outros a serem instalados.
5. Selecione as outras opções conforme necessário e selecione Create cluster (Criar cluster).

Para iniciar um cluster com o Flink a partir do AWS CLI

- Crie o cluster usando o seguinte comando:

```
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

Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

Configurando o Flink na Amazon EMR

Configurar o Flink com o Hive Metastore e o Catálogo do Glue

As EMR versões 6.9.0 e superiores da Amazon oferecem suporte ao Hive Metastore e ao AWS Glue Catalog com o conector Apache Flink para o Hive. Esta seção descreve as etapas necessárias para configurar o [Catálogo do AWS Glue](#) e o [Hive Metastore](#) com o Flink.

Tópicos

- [Usar o Hive Metastore](#)
- [Use o catálogo de dados AWS Glue](#)

Usar o Hive Metastore

1. Crie um EMR cluster com a versão 6.9.0 ou superior e pelo menos dois aplicativos: Hive e Flink.
2. Use [script runner](#) para executar o script a seguir como função de etapa:

```
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

Cancel
Save

Use o catálogo de dados AWS Glue

1. Crie um EMR cluster com a versão 6.9.0 ou superior e pelo menos dois aplicativos: Hive e Flink.
2. Selecione Usar com metadados da tabela do Hive nas configurações do Catálogo de Dados do AWS Glue para habilitar o Catálogo de Dados no cluster.
3. Use o [script runner](#) para executar o seguinte script como uma função de etapa: [Execute comandos e scripts em um EMR cluster da 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

Configurar o Flink com um arquivo de configuração

Você pode usar a EMR configuração da Amazon API para configurar o Flink com um arquivo de configuração. Os arquivos que podem ser configurados no API são:

- `flink-conf.yaml`
- `log4j.properties`
- `flink-log4j-session`
- `log4j-cli.properties`

O principal arquivo de configuração para o Flink é `flink-conf.yaml`.

Configurar o número de slots de tarefa que são usados para o Flink na AWS CLI

1. Crie um arquivo `configurations.json`, com o seguinte conteúdo:

```
[
  {
    "Classification": "flink-conf",
    "Properties": {
      "taskmanager.numberOfTaskSlots": "2"
    }
  }
]
```

]

2. Em seguida, crie um cluster com a seguinte configuração:

```
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

Você também pode alterar algumas configurações com o API Flink. Para obter mais informações, consulte [Concepts](#) na documentação do Flink.

Com a Amazon EMR versão 5.21.0 e posterior, você pode substituir as configurações do cluster e especificar classificações de configuração adicionais para cada grupo de instâncias em um cluster em execução. Você faz isso usando o EMR console da Amazon, o AWS Command Line Interface (AWS CLI) ou AWS SDK o. Para obter mais informações, consulte [Supplying a Configuration for an Instance Group in a Running Cluster](#).

Opções de paralelismo

Como proprietário da aplicação, você sabe quais recursos atribuir a tarefas no Flink. Para os exemplos nesta documentação, use o mesmo número de tarefas que as instâncias de tarefa que você usa para a aplicação. Geralmente, recomendamos isso para o nível de paralelismo inicial, mas também é possível aumentar a granularidade do paralelismo usando slots de tarefa, que geralmente não excedem o número de [núcleos virtuais](#) por instância. Para obter mais informações sobre a arquitetura do Flink, consulte [Concepts](#) na documentação do Flink.

Configurando o Flink em um EMR cluster com vários nós primários

O JobManager do Flink permanece disponível durante o processo de failover do nó primário em um EMR cluster da Amazon com vários nós primários. A partir do Amazon EMR 5.28.0, a

JobManager alta disponibilidade também é ativada automaticamente. Nenhuma configuração manual é necessária.

Com EMR as versões 5.27.0 ou anteriores da Amazon, esse JobManager é um único ponto de falha. Quando o JobManager falha, ele perde todos os estados de trabalho e não retoma os trabalhos em execução. Você pode ativar a JobManager alta disponibilidade configurando a contagem de tentativas do aplicativo, marcando o ponto de verificação e ativando o armazenamento ZooKeeper como estado para o Flink, conforme demonstra o exemplo a seguir:

```
[
  {
    "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"
    }
  }
]
```

Você deve configurar o máximo de tentativas principais de aplicativos YARN e tentativas de aplicativo para o Flink. Para obter mais informações, consulte [Configuração da alta disponibilidade do YARN cluster](#). Você também pode configurar o ponto de verificação do Flink para reiniciar a JobManager recuperação de trabalhos em execução de pontos de verificação concluídos anteriormente. Para obter mais informações, consulte [Flink checkpointing](#).

Configurar o tamanho do processo de memória

Para EMR versões da Amazon que usam o Flink 1.11.x, você deve configurar o tamanho total do processo de memória para JobManager (`jobmanager.memory.process.size`) e TaskManager (`taskmanager.memory.process.size`) in `flink-conf.yaml` Você pode definir esses valores configurando o cluster com a configuração API ou descomentando manualmente esses campos por meio de. SSH O Flink fornece os valores padrão a seguir.

- `jobmanager.memory.process.size`: 1600m
- `taskmanager.memory.process.size`: 1728m

Para excluir o JVM metaspesço e a sobrecarga, use o tamanho total da memória do Flink () `taskmanager.memory.flink.size` em vez de `taskmanager.memory.process.size`. O valor padrão para `taskmanager.memory.process.size` é 1280m. Não é recomendável definir `taskmanager.memory.process.size` e `taskmanager.memory.process.size`.

Todas as EMR versões da Amazon que usam o Flink 1.12.0 e posteriores têm os valores padrão listados no conjunto de código aberto do Flink como valores padrão na AmazonEMR, então você não precisa configurá-los sozinho.

Configurar o tamanho do arquivo de saída de log

Os contêineres de aplicações Flink criam e gravam em três tipos de arquivos de log: arquivos `.out`, arquivos `.log` e arquivos `.err`. Somente os arquivos `.err` são compactados e removidos do sistema de arquivos, enquanto os arquivos de log `.log` e `.out` permanecem no sistema de arquivos. Para garantir que esses arquivos de saída continuem gerenciáveis e que o cluster continue estável, é possível configurar a alternância de logs `log4j.properties` para definir um número máximo de arquivos e limitar o tamanho deles.

Amazon EMR versões 5.30.0 e posteriores

A partir do Amazon EMR 5.30.0, o Flink usa a estrutura de registro `log4j2` com o nome de classificação de configuração. O exemplo de configuração a seguir demonstra `flink-log4j`. o formato `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 versões 5.29.0 e anteriores

Com EMR as versões 5.29.0 e anteriores da Amazon, o Flink usa a estrutura de registro log4j. O exemplo de configuração a seguir demonstra o formato 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"
    },
  }
]

```

Configurar o Flink para ser executado com o Java 11

As EMR versões 6.12.0 e superiores da Amazon oferecem suporte ao tempo de execução do Java 11 para o Flink. As seções a seguir descrevem como configurar o cluster para fornecer suporte ao runtime do Java 11 para o Flink.

Tópicos

- [Configurar o Flink para Java 11 ao criar um cluster](#)
- [Configurar o Flink para Java 11 em um cluster em execução](#)
- [Confirmar o runtime do Java para o Flink em um cluster em execução](#)

Configurar o Flink para Java 11 ao criar um cluster

Use as etapas a seguir para criar um EMR cluster com o tempo de execução do Flink e do Java 11. O arquivo de configuração ao qual você adiciona suporte ao runtime do Java 11 é `flink-conf.yaml`.

Console

Para criar um cluster com o tempo de execução do Flink e do Java 11 no console

1. Faça login no AWS Management Console e abra o EMR console da Amazon em <https://console.aws.amazon.com/emr>.
2. Escolha Clusters EMR em EC2 Ativado no painel de navegação e, em seguida, Criar cluster.
3. Selecione a EMR versão 6.12.0 ou superior da Amazon e escolha instalar o aplicativo Flink. Selecione qualquer outra aplicação que você queira instalar no cluster.
4. Continue configurando o cluster. Na seção opcional Configurações de software, use a opção padrão Inserir configuração e insira a seguinte configuração:

```
[
  {
    "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. Continue configurando e iniciando o cluster.

AWS CLI

Para criar um cluster com o tempo de execução do Flink e do Java 11 a partir do CLI

1. Crie um arquivo de configuração `configurations.json` que configure o Flink para usar o 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. A partir do AWS CLI, crie um novo EMR cluster com a EMR versão 6.12.0 ou superior da Amazon e instale o aplicativo Flink, conforme mostrado no exemplo a seguir:

```
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
```

Configurar o Flink para Java 11 em um cluster em execução

Use as etapas a seguir para atualizar um EMR cluster em execução com o Flink e o Java 11 runtime. O arquivo de configuração ao qual você adiciona suporte ao runtime do Java 11 é `flink-conf.yaml`.

Console

Para atualizar um cluster em execução com o Flink e o tempo de execução do Java 11 no console

1. Faça login no AWS Management Console e abra o EMR console da Amazon em <https://console.aws.amazon.com/emr>.
2. Escolha Clusters EMR EC2 em Ativado no painel de navegação e, em seguida, selecione o cluster que você deseja atualizar.

Note

O cluster deve usar a EMR versão 6.12.0 ou superior da Amazon para oferecer suporte ao Java 11.

3. Selecione a guia Configuração.
4. Na seção Configurações do grupo de instâncias, selecione o grupo de instâncias Em execução que você deseja atualizar e escolha Reconfigurar no menu de ações da lista.
5. Reconfigure o grupo de instâncias com a opção Editar atributos, conforme mostrado a seguir. Selecione Adicionar nova configuração após cada.

Classificação	Propriedade	Valor
flink-conf	containerized.task manager.env.JAVA_H OME	/usr/lib/jvm/jre-1 1
flink-conf	containerized.mast er.env.JAVA_HOME	/usr/lib/jvm/jre-1 1
flink-conf	env.java.home	/usr/lib/jvm/jre-1 1

6. Selecione Salvar alterações para adicionar as configurações.

AWS CLI

Para atualizar um cluster em execução para usar o tempo de execução do Flink e do Java 11 a partir do CLI

Você pode usar o comando `modify-instance-groups` para especificar configurações para cada grupo de instâncias em um cluster em execução.

1. Primeiro, crie um arquivo de configuração `configurations.json` que configure o Flink para usar o Java 11. No exemplo a seguir, substitua `ig-1xxxxxxx9` com o ID do grupo de instâncias que você deseja reconfigurar. Salve o arquivo no mesmo diretório em que você executará o comando `modify-instance-groups`.

```
[
  {
    "InstanceGroupId": "ig-1xxxxxxxx9",
    "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. A partir do AWS CLI, execute o seguinte comando. Substitua o ID do grupo de instâncias que você deseja reconfigurar:

```
aws emr modify-instance-groups --cluster-id j-2AL4XXXXXX5T9 \
--instance-groups file://configurations.json
```

Confirmar o runtime do Java para o Flink em um cluster em execução

Para determinar o tempo de execução do Java para um cluster em execução, faça login no nó primário com SSH conforme descrito em [Conecte-se ao nó primário com SSH](#). Em seguida, execute o seguinte comando:

```
ps -ef | grep flink
```

O comando `ps` com a opção `-ef` lista todos os processos que estão em execução no sistema. É possível filtrar essa saída com `grep` para encontrar menções à string `flink`. Revise a saída do valor do Java Runtime Environment (JRE), `jre-XX`. Na saída a seguir, `jre-11` indica que o Java 11 está selecionado em runtime para o 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.
```

Como alternativa, [faça login no nó primário com SSH](#) e inicie uma YARN sessão do Flink com o comando `flink-yarn-session -d`. A saída mostra a Java Virtual Machine (JVM) para Flink, `java-11-amazon-corretto` no exemplo a seguir:

```
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
```

Usar trabalhos do Flink no Amazon EMR

Há várias maneiras de interagir com o Flink no Amazon EMR: por meio do console, da interface do Flink encontrada na interface de usuário de rastreamento e ResourceManager na linha de comando. Você pode enviar um arquivo JAR para uma aplicação Flink com qualquer uma destas opções. Depois de enviar um arquivo JAR, ele se torna um trabalho gerenciado pelo Flink JobManager. O JobManager está localizado no nó YARN que hospeda o daemon do Application Master da sessão Flink.

Você pode executar uma aplicação do Flink como um trabalho do YARN em um cluster de execução prolongada ou em um cluster transitório. Em um trabalho de execução prolongada, você pode enviar vários trabalhos do Flink para um cluster do Flink em execução no Amazon EMR. Se você executar um trabalho do Flink em um cluster transitório, seu cluster do Amazon EMR existirá somente pelo tempo necessário para executar a aplicação Flink; portanto, você pagará somente pelos recursos e pelo tempo usados. Você pode enviar um trabalho do Flink usando a operação de API `AddSteps` do Amazon EMR como um argumento de etapa para a operação `RunJobFlow` ou usando os comandos AWS CLI, `add-steps` ou `create-cluster`.

Iniciar uma aplicação YARN do Flink como uma etapa em um cluster de execução prolongada

Para iniciar uma aplicação Flink para a qual vários clientes possam enviar trabalhos por meio de operações de API do YARN, é necessário criar um cluster ou adicionar uma aplicação Flink a um

cluster já existente. Para obter instruções sobre como criar um novo cluster, consulte [Criar um cluster com o Flink](#). Para iniciar uma sessão do YARN em um cluster atual, use as etapas a seguir no console, na AWS CLI ou no SDK do Java.

Note

O comando `flink-yarn-session` foi incluído no Amazon EMR versão 5.5.0 como um wrapper para o script `yarn-session.sh` para simplificar a execução. Se você usa uma versão anterior do Amazon EMR, substitua `bash -c "/usr/lib/flink/bin/yarn-session.sh -d"` por `Arguments` no console ou `Args` no comando AWS CLI.

Enviar um trabalho do Flink que está em um cluster existente no console

Envie a sessão do Flink usando o comando `flink-yarn-session` em um cluster já existente.

1. [Abra o console do Amazon EMR em https://console.aws.amazon.com/emr](https://console.aws.amazon.com/emr).
2. Na lista de clusters, selecione o cluster que você iniciou anteriormente.
3. Na página de detalhes do cluster, selecione Steps (Etapas), Add Step (Adicionar etapa).
4. Use as diretrizes a seguir para inserir os parâmetros e escolha Adicionar.

Parâmetro	Descrição
Step type (Tipo de etapa)	JAR personalizado
Nome	Um nome para ajudar você a identificar a etapa. Por exemplo, <code>< example-flink-step-name ></code> .
Jar location (Local de jar)	command-runner.jar
Arguments (Argumentos)	O comando <code>flink-yarn-session</code> com argumentos adequados para seu aplicativo. Por exemplo, <code>flink-yarn-session -d</code> inicia uma sessão do Flink em seu cluster YARN em um estado separado (). -d Consulte a configuração do YARN na

Parâmetro	Descrição
	documentação do Flink mais recente para obter mais detalhes sobre argumentos.

Enviar um trabalho do Flink que está em um cluster existente com a AWS CLI

- Use o comando `add-steps` para adicionar um trabalho do Flink a um cluster de execução prolongada. O comando de exemplo a seguir especifica `Args="flink-yarn-session", "-d"` para iniciar uma sessão do Flink em seu cluster do YARN em um estado desconectado (`-d`). Consulte a [configuração do YARN](#) na documentação do Flink mais recente para obter mais detalhes sobre argumentos.

```
aws emr add-steps --cluster-id <j-XXXXXXX> --steps Type=CUSTOM_JAR,Name=<example-flink-step-name>,Jar=command-runner.jar,Args="flink-yarn-session","-d"
```

Enviar o trabalho para uma aplicação Flink existente em um cluster de execução prolongada

Caso já tenha uma aplicação Flink em um cluster de execução prolongada, você poderá especificar o ID da aplicação Flink do cluster para enviar trabalhos para ele. Para obter o ID do aplicativo, execute `yarn application -list` na operação da [YarnClient](#) API AWS CLI ou por meio dela:

```
$ yarn application -list
16/09/07 19:32:13 INFO client.RMPProxy: 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
```

O ID da aplicação para esta sessão do Flink é `application_1473169569237_0002`, que pode ser usado para enviar trabalhos para a aplicação pela AWS CLI ou por um SDK.

Example SDK para 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));
```

Example 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>
```

Enviar um trabalho transitório do Flink

Os exemplos a seguir iniciam um cluster transitório que executa um trabalho do Flink e é terminado na conclusão.

Example SDK para 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());

}

}

```

Example AWS CLI

Use o subcomando `create-cluster` para criar um cluster transitório que termina quando o trabalho do Flink é concluído:

```

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/"

```

Usar o shell Scala

O shell Scala do Flink para clusters do EMR está configurado somente para iniciar novas sessões do YARN. Você pode usar o shell Scala seguindo o procedimento abaixo.

Usar o shell Flink Scala no nó primário

1. Faça login no nó primário com SSH, conforme descrito em [Connect to the primary node with SSH](#).
2. Digite o seguinte para iniciar um shell:

No Amazon EMR 5.5.0 e versões posteriores, você pode usar o comando a seguir para iniciar um cluster do Yarn para o shell Scala com um TaskManager.

```
% flink-scala-shell yarn 1
```

Em versões anteriores do Amazon EMR, use:

```
% /usr/lib/flink/bin/start-scala-shell.sh yarn 1
```

Isso inicia o shell Scala do Flink para que você possa usar o Flink interativamente. Assim como com outras interfaces e opções, você pode escalar o valor da opção `-n` usado no exemplo com base no número de tarefas que você deseja executar a partir do shell.

Para obter mais informações, consulte [Scala REPL](#) na documentação oficial do Apache Flink.

Localizar a interface Web do Flink

O Application Master que pertence à aplicação Flink hospeda a interface da Web do Flink. É uma forma alternativa de enviar um JAR como trabalho ou de visualizar o status atual de outros trabalhos. A interface da web do Flink ficará ativa desde que você tenha um sessão do Flink em execução. Se você tiver um trabalho do YARN de execução prolongada já ativo, siga as instruções no tópico [Connect to the primary node with SSH](#) no Guia de gerenciamento do Amazon EMR para conectar-se ao YARN ResourceManager. Por exemplo, se você tiver configurado um túnel SSH e ativado um proxy em seu navegador, será possível escolher a conexão do ResourceManager em Conexões na página de detalhes do cluster do EMR.

Cluster: Development Cluster **Waiting** Cluster ready after last step completed.

Connections:  [Resource Manager ... \(View All\)](#)

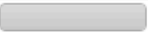
Depois de encontrar o ResourceManager, selecione a aplicação YARN que está hospedando uma sessão do Flink. Selecione o link na coluna Tracking UI.

Lo

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


Na interface da web do Flink, você pode visualizar a configuração, enviar seu próprio JAR personalizado como um trabalho ou monitorar os trabalhos em andamento.


 Apache Flink Dashboard

- Overview
- Running Jobs
- Completed Jobs
- Task Managers
- Job Manager
- Submit new Job

Overview
Version: 1.1.1

 1
Task Managers

 1
Task Slots

 1
Available Task Slots

Total Jobs	
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

Escalador automático do Flink

Visão geral

As versões 6.15.0 e superiores do Amazon EMR oferecem suporte ao escalador automático do Flink. A funcionalidade de escalador automático de trabalhos coleta métricas da execução de trabalhos de streaming do Flink e escala automaticamente os vértices individuais do trabalho. Isso reduz a contrapressão e satisfaz a meta de utilização definida por você.

Para obter mais informações, consulte a seção [Autoscaler](#) da documentação do Apache Flink Kubernetes Operator.

Considerações

- O escalador automático do Flink é compatível com o Amazon EMR 6.15.0 e versões superiores.
- O escalador automático do Flink é compatível somente com trabalhos de streaming.
- Só há suporte para o agendador adaptável. Não há suporte para o agendador padrão.
- Recomendamos habilitar o ajuste de escala de clusters para permitir o provisionamento dinâmico de recursos. O Ajuste de Escala Gerenciado do Amazon EMR é preferencial porque a avaliação da métrica ocorre a cada cinco a dez segundos. Nesse intervalo, seu cluster pode se ajustar mais prontamente à mudança nos recursos necessários do cluster.

Habilitação do escalador automático

Use as etapas a seguir para habilitar o escalador automático do Flink ao criar um cluster do Amazon EMR no EC2.

1. No console do Amazon EMR, crie um cluster do EMR.
 - a. Escolha a versão `emr-6.15.0` do Amazon EMR ou posterior. Selecione o pacote de aplicações Flink e quaisquer outras aplicações que você queira incluir no 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. Na opção Ajuste de escala e provisionamento de clusters, selecione Usar Ajuste de Escala Gerenciado do 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. Na seção Configurações de software, insira a configuração a seguir para habilitar o escalador automático do Flink. Em cenários de teste, defina o intervalo de decisão, o intervalo da janela de métricas e o intervalo de estabilização para um valor menor de modo que o trabalho tome imediatamente uma decisão de ajuste de escala para facilitar a verificação.

```
[
{
```

```
"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. Selecione ou defina quaisquer outras configurações da maneira que preferir e crie o cluster habilitado pelo escalador automático do Flink.

Configurações do escalador automático

Esta seção aborda a maioria das configurações que você pode alterar com base nas suas necessidades específicas.

Note

Com configurações baseadas em tempo, como `time`, `interval` e `window`, a unidade padrão quando nenhuma unidade é especificada é milissegundos. Portanto, um valor de 30 sem sufixo é igual a 30 milissegundos. Para outras unidades de tempo, inclua o sufixo apropriado de `s` para segundos, `m` para minutos ou `h` para horas.

Tópicos

- [Configurações de loop do escalador automático](#)
- [Agregação de métricas e configurações de histórico](#)
- [Configurações no nível do vértice do trabalho](#)
- [Configurações de processamento do backlog](#)
- [Configurações da operação de ajuste de escala](#)

Configurações de loop do escalador automático

O escalador automático busca as métricas do nível do vértice do trabalho para cada intervalo de tempo configurável, converte-as em ações de escala, estima o novo paralelismo do vértice do trabalho e o recomenda ao agendador de trabalhos. As métricas são coletadas somente após o horário de reinício do trabalho e o intervalo de estabilização do cluster.

Chave de configuração	Valor padrão	Descrição	Exemplos de valores
<code>job.autoscaler.enabled</code>	<code>false</code>	Habilita o ajuste de escala automático no cluster do Flink.	<code>true</code> , <code>false</code>
<code>job.autoscaler.decision.interval</code>	<code>60s</code>	Intervalo de decisões do escalador automático.	<code>30</code> (a unidade padrão é milissegundos), <code>5m</code> , <code>1h</code>
<code>job.autoscaler.restart.time</code>	<code>3m</code>	O tempo de reinício esperado a ser usado até que o operador possa determiná-lo de forma confiável com o histórico.	<code>30</code> (a unidade padrão é milissegundos), <code>5m</code> , <code>1h</code>
<code>job.autoscaler.stabilization.interval</code>	<code>300s</code>	O período de estabilização no qual nenhum novo ajuste de escala será executado.	<code>30</code> (a unidade padrão é milissegundos), <code>5m</code> , <code>1h</code>
<code>job.autoscaler.debug.logs.interval</code>	<code>300s</code>	Intervalo dos logs de depuração do escalador automático.	<code>30</code> (a unidade padrão é milissegundos), <code>5m</code> , <code>1h</code>

Agregação de métricas e configurações de histórico

O escalador automático busca as métricas, faz a agregação delas ao longo de uma janela deslizante com base no tempo e elas são avaliadas em decisões de ajuste de escala. O histórico de decisões

do ajuste de escala para cada vértice do trabalho é utilizado para estimar novos paralelismos. Eles têm validade baseada na idade e no tamanho do histórico (pelo menos 1).

Chave de configuração	Valor padrão	Descrição	Exemplos de valores
<code>job.autoscaler.metrics.window</code>	600s	Scaling metrics aggregation window size.	30 (a unidade padrão é milissegundos), 5m, 1h
<code>job.autoscaler.history.max.count</code>	3	Número máximo de decisões de ajuste de escala anteriores a serem retidas por vértice.	1 para Integer.MAX_VALUE
<code>job.autoscaler.history.max.age</code>	24h	Número mínimo de decisões de ajuste de escala anteriores a serem retidas por vértice.	30 (a unidade padrão é milissegundos), 5m, 1h

Configurações no nível do vértice do trabalho

O paralelismo de cada vértice do trabalho é modificado com base na utilização de destino e vinculado aos limites mínimo e máximo do paralelismo. Não é recomendável definir uma utilização de destino próxima de 100% (ou seja, valor de 1), e o limite de utilização funciona como um buffer para lidar com as flutuações intermediárias da carga.

Chave de configuração	Valor padrão	Descrição	Exemplos de valores
<code>job.autoscaler.target.utilization</code>	0.7	Utilização de destino do vértice.	0 - 1
<code>job.autoscaler.target.utilization.boundary</code>	0.4	Limite de utilização de destino do vértice. O ajuste de escala não será realizado se a taxa de	0 - 1

Chave de configuração	Valor padrão	Descrição	Exemplos de valores
		processamento atual estiver entre $[\text{target_rate} / (\text{target_utilization} - \text{boundary})$ e $(\text{target_rate} / (\text{target_utilization} + \text{boundary}))$.	
<code>job.autoscaler.vertex.min-parallelism</code>	1	O paralelismo mínimo que o escalador automático pode usar.	0 - 200
<code>job.autoscaler.vertex.max-parallelism</code>	200	O paralelismo máximo que o escalador automático pode usar. Observe que esse limite será ignorado se ele for maior que o paralelismo máximo definido na configuração do Flink ou diretamente em cada operador.	0 - 200

Configurações de processamento do backlog

O vértice do trabalho precisa de recursos extras para lidar com os eventos pendentes, ou backlogs, que se acumulam durante o período da operação de escala. Isso também é conhecido como duração de catch-up. Se o tempo de processamento do backlog exceder o valor de `lag - threshold` configurado, a utilização de destino do vértice do trabalho aumentará para o nível máximo. Isso ajuda a evitar operações desnecessárias de ajuste de escala durante o processamento do backlog.

Chave de configuração	Valor padrão	Descrição	Exemplos de valores
<code>job.autoscaler.backlog-processing.lag-threshold</code>	5m	Limite de atraso que evitará ajustes de escala desnecessários e removerá as mensagens pendentes responsáveis pelo atraso.	30 (a unidade padrão é milissegundos), 5m, 1h
<code>job.autoscaler.catch-up.duration</code>	15m	A duração de destino para o processamento completo de qualquer backlog após uma operação de ajuste de escala. Defina como 0 para desabilitar o ajuste de escala baseado em backlog.	30 (a unidade padrão é milissegundos), 5m, 1h

Configurações da operação de ajuste de escala

O escalador automático não executa a operação de redução vertical de escala imediatamente após uma operação de aumento vertical de escala dentro do período de carência. Isso evita ciclos desnecessários de operações de aumento e redução vertical de escala causados por flutuações temporárias de carga.

Podemos usar a taxa de operação de redução vertical de escala para diminuir gradualmente o paralelismo e liberar recursos que atendam a picos de carga temporários. Isso também ajuda a evitar operações secundárias desnecessárias de aumento vertical de escala após uma grande operação de redução vertical de escala.

Podemos detectar uma operação ineficaz de aumento vertical de escala com base no histórico anterior de decisões de ajuste de escala de vértices de trabalho para evitar mais alterações no paralelismo.

Chave de configuração	Valor padrão	Descrição	Exemplos de valores
<code>job.autoscaler.scale-up.grace-period</code>	1h	Duração em que nenhuma redução vertical de escala de um vértice é permitida após ele ter sido aumentado.	30 (a unidade padrão é milissegundos), 5m, 1h
<code>job.autoscaler.scale-down.max-factor</code>	0.6	Fator máximo de redução vertical de escala. Um valor de 1 significa que não há limite na redução vertical de escala; 0.6 significa que o trabalho só pode ser reduzido verticalmente com 60% do paralelismo original.	0 - 1
<code>job.autoscaler.scale-up.max-factor</code>	100000.	Taxa máxima de aumento vertical de escala. Um valor de 2.0 significa que o trabalho só pode aumentar a escala verticalmente com 200% do paralelismo atual.	0 - Integer.MAX_VALUE
<code>job.autoscaler.scaling.effectiveness.detection.enabled</code>	false	Se deve habilitar a detecção de operações ineficazes de ajuste de escala e permitir que o escalador automático o bloqueie novos aumentos verticais.	true, false

Otimização dos tempos de reinício dos trabalhos para operações de recuperação e ajuste de escala de tarefas

Quando uma tarefa falha ou ocorre uma operação de ajuste de escala, o Flink tenta reexecutar a tarefa com base no último ponto de verificação concluído. O processo de reinicialização pode levar

um minuto ou mais para ser executado, dependendo do tamanho do estado do ponto de verificação e do número de tarefas paralelas. Durante o período de reinicialização, as tarefas de backlog podem se acumular para o trabalho. No entanto, existem algumas maneiras de o Flink otimizar a velocidade de recuperação e reinicialização dos gráficos de execução para melhorar a estabilidade do trabalho.

Esta página descreve algumas maneiras que o Flink no Amazon EMR pode usar para melhorar o tempo de reinício de trabalhos durante as operações de recuperação ou ajuste de escala de tarefas.

Tópicos

- [Recuperação local de tarefas](#)
- [Ponto de verificação incremental genérico baseado em logs](#)
- [Recuperação refinada](#)
- [Mecanismo de reinício combinado no agendador adaptável](#)

Recuperação local de tarefas

Note

A recuperação local de tarefas é compatível com o Amazon EMR 6.0.0 e versões superiores.

Com os pontos de verificação do Flink, cada tarefa produz um snapshot do próprio estado que o Flink grava em um armazenamento distribuído, como o Amazon S3. Em casos de recuperação, as tarefas restauram o estado no armazenamento distribuído. O armazenamento distribuído oferece tolerância a falhas e pode redistribuir o estado durante o reajuste de escala por ser acessível a todos os nós.

No entanto, um armazenamento distribuído remoto também tem uma desvantagem: todas as tarefas devem ler seu estado de um local remoto na rede. Isso pode resultar em tempos de recuperação longos para estados grandes durante operações de recuperação ou ajuste de escala de tarefas.

Esse problema de tempo de recuperação longo é resolvido pela recuperação local de tarefas. As tarefas gravam seu estado no ponto de verificação em um armazenamento secundário local para a tarefa, como em um disco local. Elas também armazenam seu estado no armazenamento principal, ou no Amazon S3, no nosso caso. Durante a recuperação, o agendador programa as tarefas no mesmo gerenciador de tarefas em que elas foram executadas anteriormente, para que possam se

recuperar do armazenamento de estado local em vez de ler do armazenamento de estado remoto. Para obter mais informações, consulte [Task-Local Recovery](#) na Documentação do Apache Flink.

Nossos testes de benchmark com exemplos de trabalhos mostraram que o tempo de recuperação foi reduzido de minutos para alguns segundos com a recuperação local da tarefa ativada.

Para habilitar a recuperação local de tarefas, defina as configurações a seguir no seu arquivo `flink-conf.yaml`. Especifique o valor do intervalo do ponto de verificação em milissegundos.

```
state.backend.local-recovery: true
state.backend: hasmap or rocksdb
state.checkpoints.dir: s3://storage-location-bucket-path/checkpoint
execution.checkpointing.interval: 15000
```

Ponto de verificação incremental genérico baseado em logs

Note

Os pontos de verificação incrementais genéricos baseados em logs são compatíveis com o Amazon EMR 6.10.0 e versões superiores.

O recurso de ponto de verificação incremental genérico baseado em logs foi adicionado ao Flink 1.16 para melhorar a velocidade dos pontos de verificação. Um intervalo de ponto de verificação mais rápido costuma resultar em uma redução do trabalho de recuperação porque menos eventos precisam ser reprocessados após a recuperação. Para obter mais informações, consulte [Improving speed and stability of checkpointing with generic log-based incremental checkpoints](#) no blog do Apache Flink.

Com exemplos de trabalhos, nossos testes de comparação mostraram que o tempo do ponto de verificação foi reduzido de minutos para alguns segundos com o ponto de verificação incremental genérico baseado em logs.

Para habilitar os pontos de verificação incrementais genéricos baseados em logs, defina as configurações a seguir no seu arquivo `flink-conf.yaml`. Especifique o valor do intervalo do ponto de verificação em milissegundos.

```
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
```

Recuperação refinada

Note

O suporte à recuperação refinada do agendador padrão está disponível no Amazon EMR 6.0.0 e versões superiores. O suporte à recuperação refinada no agendador adaptável está disponível no Amazon EMR 6.15.0 e versões superiores.

Quando uma tarefa falha durante a execução, o Flink redefine todo o gráfico de execução e aciona a reexecução completa com base no último ponto de verificação concluído. Isso é mais caro do que apenas reexecutar as tarefas que falharam. A recuperação refinada reinicia somente o componente conectado ao pipeline da tarefa que falhou. No exemplo a seguir, o gráfico do trabalho tem cinco vértices (de A a E). É feito o pipeline de todas as conexões entre os vértices com distribuição pontual, e o `parallelism.default` do trabalho é definido como 2.

```
A # B # C # D # E
```

Neste exemplo, há um total de dez tarefas em execução. O primeiro pipeline (de a1 a e1) é executado em um TaskManager (TM1) e o segundo pipeline (de a2 a e2) é executado em outro TaskManager (TM2).

```
a1 # b1 # c1 # d1 # e1
a2 # b2 # c2 # d2 # e2
```

Há dois componentes conectados por pipeline: a1 # e1 e a2 # e2. Se TM1 ou TM2 falhar, a falha afetará somente as cinco tarefas no pipeline em que TaskManager estava em execução. A estratégia de reinicialização só inicia o componente do pipeline afetado.

A recuperação refinada funciona somente com trabalhos perfeitamente paralelos do Flink. Não é compatível com operações de `keyBy()` ou `redistribute()`. Para obter mais informações, consulte [FLIP-1: Fine Grained Recovery from Task Failures](#) no projeto Jira Flink Improvement Proposal.

Para habilitar a recuperação refinada, defina as configurações a seguir no seu arquivo `flink-conf.yaml`.

```
jobmanager.execution.failover-strategy: region
restart-strategy: exponential-delay or fixed-delay
```

Mecanismo de reinício combinado no agendador adaptável

Note

O mecanismo de reinício combinado no agendador adaptável é compatível com o Amazon EMR 6.15.0 e versões superiores.

O agendador adaptável pode ajustar o paralelismo do trabalho com base nos slots disponíveis. Ele reduz automaticamente o paralelismo se não houver slots suficientes disponíveis para atender ao paralelismo do trabalho configurado. Se novos slots ficarem disponíveis, o trabalho aumentará a escala verticalmente mais uma vez para o paralelismo do trabalho configurado. Um agendador adaptável evita o tempo de inatividade no trabalho quando não há recursos suficientes disponíveis. Esse é o agendador compatível com o escalador automático do Flink. Recomendamos o agendador adaptável com o Flink no Amazon EMR por esses motivos. No entanto, os agendadores adaptáveis podem fazer várias reinicializações em um curto período; uma reinicialização para cada novo recurso adicionado. Isso pode levar a uma queda na performance do trabalho.

Com o Amazon EMR 6.15.0 e versões superiores, o Flink tem um mecanismo de reinício combinado no agendador adaptável que abre uma janela de reinicialização quando o primeiro recurso é adicionado e aguarda pelo intervalo configurado da janela do padrão de um minuto. Ele executa uma única reinicialização quando há recursos suficientes disponíveis para executar o trabalho com o paralelismo configurado ou quando o intervalo expira.

Com os exemplos de trabalhos, nossos testes de comparação mostraram que esse recurso processa 10% a mais dos registros do que o comportamento padrão ao usar o agendador adaptável e o escalador automático do Flink.

Para habilitar o mecanismo de reinício combinado, defina as configurações a seguir no seu arquivo `flink-conf.yaml`.

```
jobmanager.adaptive-scheduler.combined-restart.enabled: true
jobmanager.adaptive-scheduler.combined-restart.window-interval: 1m
```

Usar trabalhos do Flink pelo Zeppelin no Amazon EMR

Introdução

O Amazon EMR 6.10.0 e versões posteriores oferecem suporte à integração de [Apache Zeppelin](#) com o Apache Flink. É possível enviar trabalhos do Flink de forma interativa por meio dos cadernos Zeppelin. Com o intérprete Flink, é possível executar consultas do Flink, definir trabalhos em lote e de transmissão do Flink e visualizar a saída nos cadernos Zeppelin. O intérprete Flink foi desenvolvido com base na API REST do Flink. Isso possibilita acessar e manipular trabalhos do Flink de dentro do ambiente Zeppelin para realizar processamento e análise de dados em tempo real.

Há quatro subintérpretes no intérprete Flink. Eles servem a outras finalidades, mas estão todos na JVM e compartilham os mesmos pontos de entrada pré-configurados do Flink (`ExecutionEnvironment`, `StreamExecutionEnvironment`, `BatchTableEnvironment`, `StreamTableEnvironment`). Os intérpretes são estes:

- `%flink`: cria `ExecutionEnvironment`, `StreamExecutionEnvironment`, `BatchTableEnvironment`, `StreamTableEnvironment` e fornece um ambiente Scala
- `%flink.pyflink`: fornece um ambiente Python
- `%flink.ssql`: fornece um ambiente SQL de transmissão
- `%flink.bsql`: fornece um ambiente SQL em lote

Pré-requisitos

- A integração do Zeppelin com o Flink é compatível com clusters criados com o Amazon EMR 6.10.0 e versões posteriores.
- Para visualizar as interfaces Web hospedadas em clusters do EMR conforme necessário para essas etapas, é necessário configurar um túnel SSH para permitir o acesso de entrada. Para obter mais informações, consulte [Configure proxy settings to view websites hosted on the primary node](#).

Configurar o Zeppelin-Flink em um cluster do EMR

Realize as etapas a seguir para configurar o Apache Flink no Apache Zeppelin para ser executado em um cluster do EMR:

1. Crie um novo cluster pelo console do Amazon EMR. Selecione a versão emr-6.10.0 ou posterior do Amazon EMR. Em seguida, escolha personalizar seu pacote de aplicações com a opção Personalizado. Inclua, no mínimo, Flink, Hadoop e Zeppelin no pacote.

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

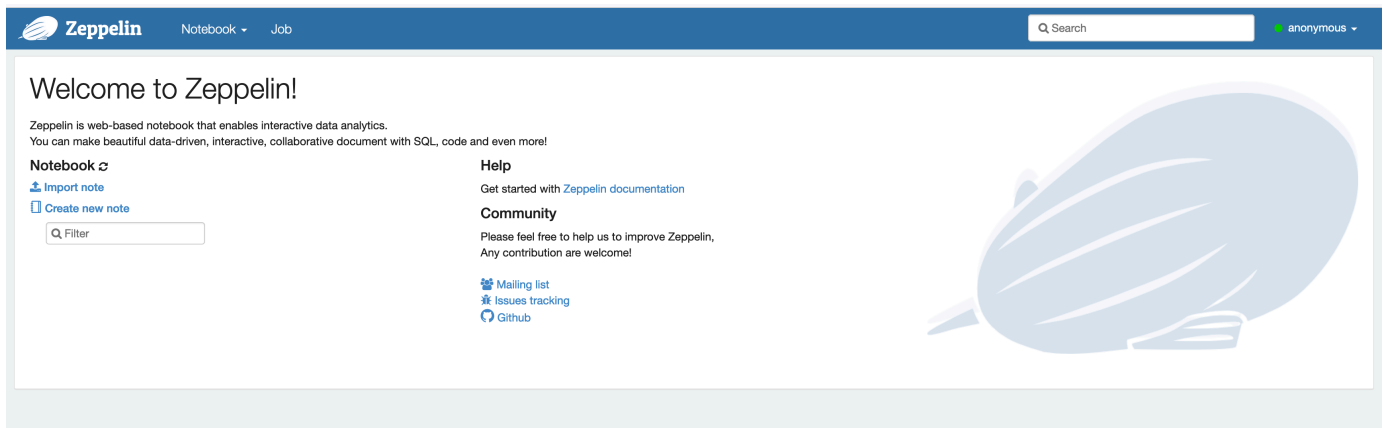
Custom

▼ Customize your application bundle

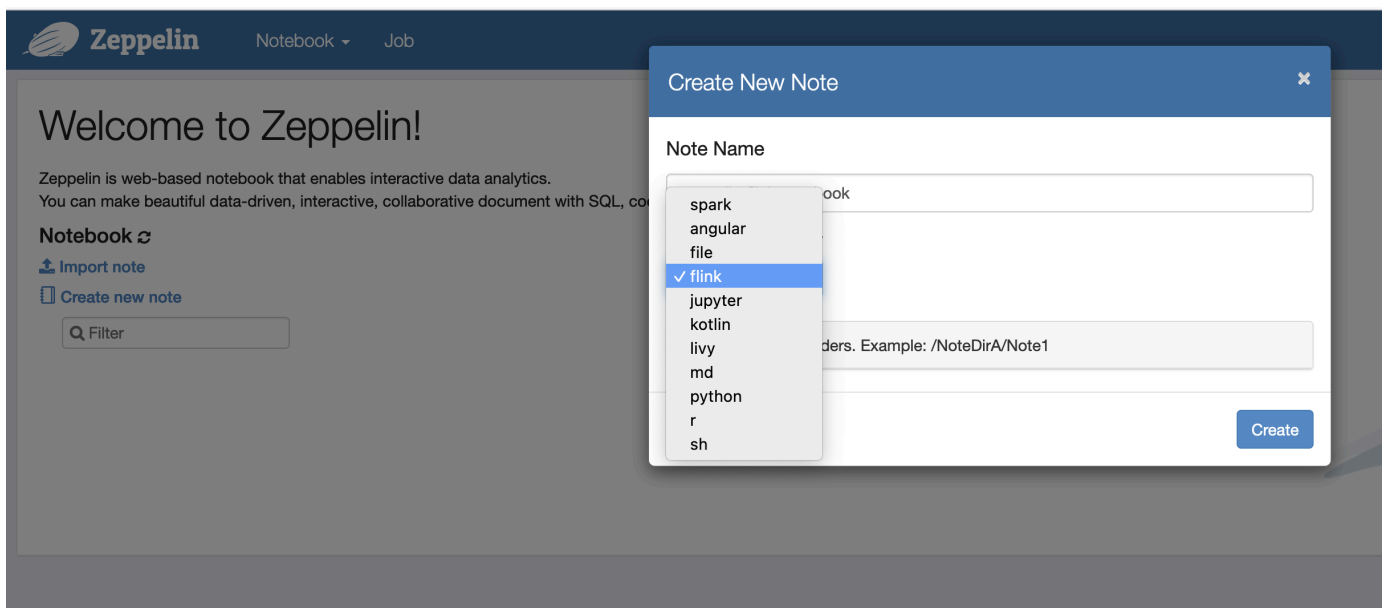
Applications included in bundle

<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. Crie o resto do cluster com as configurações de sua preferência.
3. Quando o cluster estiver em execução, selecione o cluster no console para visualizar os detalhes e abra a guia Aplicações. Selecione Zeppelin na seção Interfaces de usuário da aplicação para abrir a interface da Web do Zeppelin. Configure o acesso à interface da Web do Zeppelin com um túnel SSH para o nó primário e uma conexão proxy, conforme descrito em [Pré-requisitos](#).



4. Agora, você pode criar uma nova nota em um caderno do Zeppelin usando o Flink como intérprete padrão.



5. Consulte os exemplos de código a seguir que demonstram como executar trabalhos do Flink em um caderno do Zeppelin.

Executar trabalhos do Flink com o Zeppelin-Flink em um cluster do EMR

- Exemplo 1, Flink Scala

a) WordCount Exemplo de Batch (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 Exemplo de streaming (SCALA)

```
%flink

val data = env.fromElements("hello world", "hello flink", "hello hadoop")
data.flatMap(line => line.split("\\s"))
    .map(w => (w, 1))
    .keyBy(0)
    .sum(1)
    .print

env.execute()
```

The image displays two screenshots of Flink job execution logs. The left screenshot, titled "Batch WordCount", shows a Scala code snippet for a batch job. The code reads: `%flink`, `val data = env.fromElements("hello world", "hello flink", "hello hadoop")`, `data.flatMap(line => line.split("\\s"))`, `.map(w => (w, 1))`, `.groupBy(0)`, `.sum(1)`, `.print()`. The output shows: `data: org.apache.flink.api.scala.DataSet[String] = org.apache.flink.api.scala.DataSet@22fe7dd5`, `(flink,1)`, `(hadoop,1)`, `(hello,3)`, `(world,1)`. The right screenshot, titled "Streaming WordCount", shows the same code snippet but with `env.execute()` at the end. The output shows: `data: org.apache.flink.streaming.api.scala.DataStream[String] = org.apache.flink.streaming.api.scala.DataStream@202340f2`, a warning about deprecation, `res2: org.apache.flink.streaming.api.datastream.DataStreamSink[(String, Int)] = org.apache.flink.streaming.api.d atastream.DataStreamSink@7a0e7b2f`, and `res3: org.apache.flink.api.common.JobExecutionResult = Program execution finished`. It also states: "Job with JobID a78d1a9b33c0fb71b8d53f00cda51030 has finished." and "Job Runtime: 8485 ms".

- Exemplo 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;
```

The screenshot shows a PyFlink job execution interface. At the top, the SQL code is displayed:

```
%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;
```

Below the code, a table displays the results of the query:

id	data
3	Hello
3	World
3	Hello
3	World
4	ADD
4	ADD
5	LINE
5	LINE

At the bottom left, a status message reads: "Took 20 sec. Last updated by anonymous at May 04 2023, 3:07:03 PM. (outdated)"

- Exemplo 3, Pyflink. Observe que você deve carregar seu próprio arquivo de texto de amostra nomeado `word.txt` para seu bucket do 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. Escolha TRABALHO DO FLINK na interface do usuário do Zeppelin para acessar e visualizar a interface do usuário Web do Flink.



The screenshot shows the Zeppelin interface for a Flink job titled "Batch WordCount". The job status is "FINISHED". A red box highlights the "FLINK JOB" button, and a red arrow points to it. The job's code is displayed in a text area, and the output is shown below it. The output consists of four lines: (flink,1), (hadoop,1), (hello,3), and (world,1). The job took 56 seconds to complete and was last updated by an anonymous user on May 04, 2023, at 2:19:24 PM.

```

%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. Escolher TRABALHO DO FLINK encaminhará o console Web do Flink para outra guia do navegador.

The screenshot shows the Apache Flink Dashboard interface. On the left is a navigation sidebar with options: Overview, Jobs (Running Jobs, Completed Jobs), Task Managers, Job Manager, and Submit New Job. The main content area displays:

- Available Task Slots:** 0. Total Task Slots: 0, Task Managers: 0.
- Running Jobs:** 0. Finished: 2, Canceled: 0, Failed: 0.
- Running Job List:** A table with columns: Job Name, Start Time, Duration, End Time, Tasks, Status. It shows "No Data".
- Completed Job List:** A table with columns: Job Name, Start Time, Duration, End Time, Tasks, Status.

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

Histórico de versões do Flink

A tabela a seguir lista a versão do Flink incluída em cada versão de lançamento da AmazonEMR, junto com os componentes instalados com o aplicativo. Para obter as versões dos componentes em cada versão, consulte a seção Versão de componentes da sua versão em [Versões de lançamento do Amazon EMR 7.x](#), [Versões de lançamento do Amazon EMR 6.x](#) ou [Versões de lançamento do Amazon EMR 5.x](#).

Informações de versão do Flink

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o 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-resourcemanager, hadoop-yarn-timeline-server, flink-client,

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o Flink
		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-httfs-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-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o 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-ya rn-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-ya rn-timeline-server, flink-client, flink-jobmanager-config, hudi, delta-standalone-connectors

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o Flink
emr-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-ya rn-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-ya rn-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-ya rn-timeline-server, flink-client

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o Flink
emr-5.21.2	1.7.0	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-5.21.1	1.7.0	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-5.21.0	1.7.0	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

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Flink	Componentes instalados com o 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

Notas de lançamento do Flink por versão

Consulte as seções a seguir para ver as notas de lançamento completas.

Amazon EMR 7.2.0 - Notas de lançamento do Flink

Tipo	Descrição
Melhoria	Support a adição de rótulos personalizados ao serviço Kubernetes por tarefa do Flink por meio da configuração. <code>kubernetes.service.labels</code>

Ganglia

Note

A última versão do Amazon EMR a incluir o Ganglia foi o Amazon EMR 6.15.0. Para monitorar seu cluster, versões superiores à 6.15.0 incluem o agente [Amazon CloudWatch](#).

O projeto de código-fonte aberto Ganglia é um sistema distribuído e escalável, projetado para monitorar clusters e grades e, ao mesmo tempo, minimizar o impacto sobre o seu desempenho. Quando você habilita o Ganglia no seu cluster, pode gerar relatórios e visualizar o desempenho do cluster como um todo, bem como inspecionar o desempenho de instâncias de nós individuais. O Ganglia também é configurado para analisar e visualizar as métricas do Hadoop e do Spark. Para obter mais informações sobre o projeto de código-fonte aberto Ganglia, acesse <http://ganglia.info/>.

Ao visualizar a interface do usuário da Web do Ganglia em um navegador, você tem uma visão geral da performance do cluster, com grafos detalhando a carga, o uso da memória, a utilização da CPU e o tráfego de rede do cluster. Abaixo das estatísticas do cluster estão gráficos para cada servidor do cluster.

A tabela a seguir lista a versão do Ganglia incluída na versão mais recente da série 6.x do Amazon EMR, além dos componentes que o Amazon EMR instala com o Ganglia.

Para obter a versão dos componentes instalados com o Ganglia nessa versão, consulte [Release 6.15.0 Component Versions](#).

Informações de versão do Ganglia para o emr-6.15.0

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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-server, hadoop-yarn-nodema

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o Ganglia
		nager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, webserver

A tabela a seguir lista a versão do Ganglia incluída na versão mais recente da série 5.x do Amazon EMR, além dos componentes que o Amazon EMR instala com o Ganglia.

Para a versão dos componentes instalados com o Ganglia nesta versão, consulte Versões de componentes da [versão 5.36.2](#).

Informações sobre a versão do Ganglia para o emr-5.36.2

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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-resourcemanager, hadoop-yarn-timeline-server, webserver

Tópicos

- [Criar um cluster com o Ganglia](#)
- [Visualizar métricas do Ganglia](#)
- [Métricas do Hadoop e do Spark no Ganglia](#)
- [Histórico de versões do Ganglia](#)

Criar um cluster com o Ganglia

Note

A última versão do Amazon EMR a incluir o Ganglia foi o Amazon EMR 6.15.0. Para monitorar seu cluster, versões superiores à 6.15.0 incluem o agente [Amazon CloudWatch](#).

Para criar um cluster com o Ganglia usando o console

1. Navegue até o novo console do Amazon EMR e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte [Usar o console antigo](#).
2. Selecione Criar cluster.
3. Em Software configuration (Configuração do software), escolha All Applications (Todos os aplicativos), Core Hadoop ou Spark.
4. Prossiga com a criação do cluster usando as configurações apropriadas.

Para adicionar o Ganglia a um cluster usando o AWS CLI

No AWS CLI, você pode adicionar o Ganglia a um cluster usando `create-cluster` com o `--applications` parâmetro. Se você especificar apenas o Ganglia usando o parâmetro `--applications`, o Ganglia será o único aplicativo instalado.

- Digite o seguinte comando para adicionar o Ganglia quando você criar um cluster e substitua *myKey* pelo nome do seu par de chaves do EC2.

Note

Os caracteres de continuação de linha do Linux (`\`) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (`^`).

```
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
```

Quando você especifica a contagem de instâncias sem usar o parâmetro `--instance-groups`, um único nó principal é executado, e as instâncias restantes são executadas como nós core. Todos os nós usam o tipo de instância especificado no comando.

Note

Se você não tiver criado o perfil de serviço padrão do EMR e o perfil de instância do EC2, digite `aws emr create-default-roles` para criá-los antes de digitar o subcomando `create-cluster`.

Para obter mais informações sobre o uso dos comandos do Amazon EMR no AWS CLI, consulte <https://docs.aws.amazon.com/cli/latest/reference/emr>

Visualizar métricas do Ganglia

Note

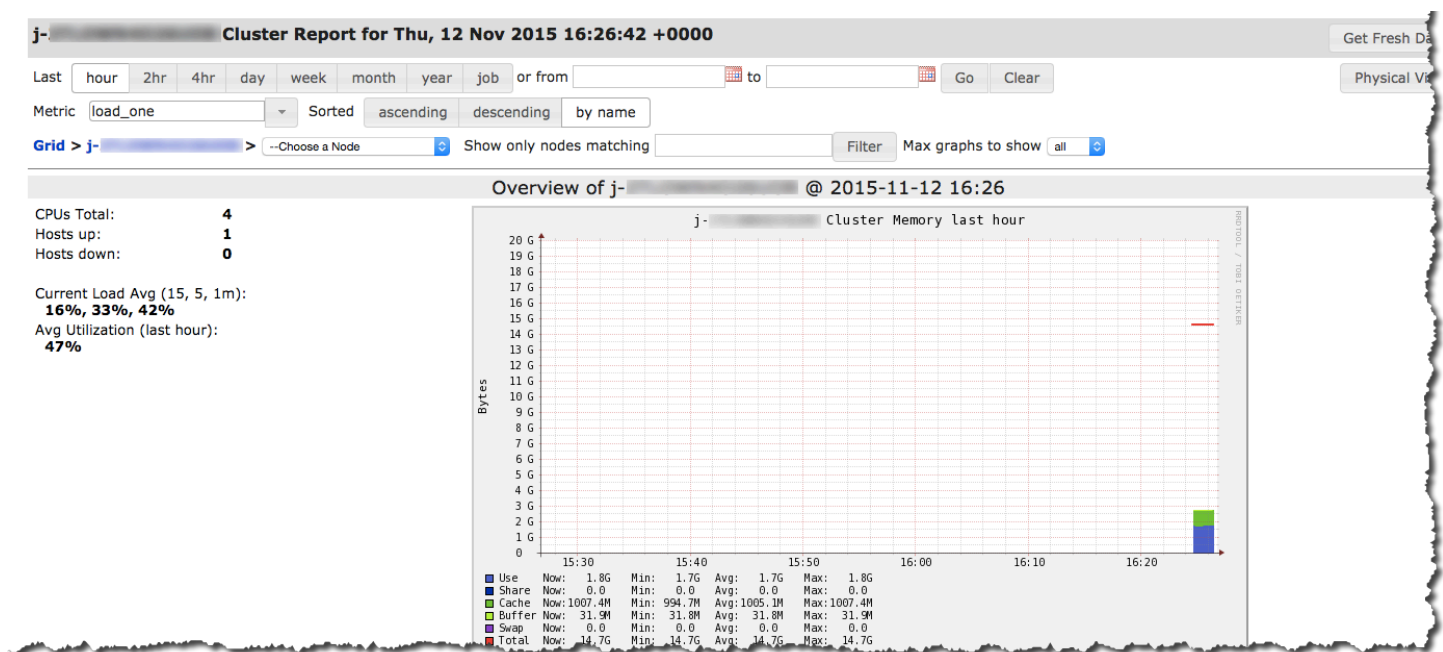
A última versão do Amazon EMR a incluir o Ganglia foi o Amazon EMR 6.15.0. Para monitorar seu cluster, versões superiores à 6.15.0 incluem o agente [Amazon CloudWatch](#).

O Ganglia fornece uma interface de usuário na web que você pode usar para visualizar as métricas que ele coleta. Quando você executa o Ganglia no Amazon EMR, a interface Web é executada no nó principal e pode ser visualizada por meio do encaminhamento de porta, também conhecido como criação de um túnel de SSH. Para obter mais informações sobre a visualização de interfaces Web no Amazon EMR, consulte [View web interfaces hosted on EMR clusters](#) no Guia de gerenciamento do Amazon EMR.

Para visualizar a interface da web do Ganglia

1. Use o SSH para formar um túnel até o nó principal e criar uma conexão segura. Para obter informações sobre como criar um túnel SSH até o nó principal, consulte a [Option 2, part 1: Set](#)

- [up an SSH tunnel to the master node using dynamic port forwarding](#) no Guia de gerenciamento do Amazon EMR.
- Instale um navegador da Web com uma ferramenta de proxy, como o FoxyProxy plug-in do Firefox, para criar um proxy SOCKS para domínios do tipo `*ec2*.amazonaws.com*`. Para obter mais informações, consulte [Option 2, part 2: Configure proxy settings to view websites hosted on the master node](#) no Guia de gerenciamento do Amazon EMR.
 - Com o proxy definido e a conexão SSH aberta, você pode visualizar a interface do usuário do Ganglia abrindo uma janela do navegador com `http://master-public-dns-name/ganglia/`, onde *master-public-dns-name* está o endereço DNS público do servidor mestre no cluster do EMR.



Métricas do Hadoop e do Spark no Ganglia

Note

A última versão do Amazon EMR a incluir o Ganglia foi o Amazon EMR 6.15.0. Para monitorar seu cluster, versões superiores à 6.15.0 incluem o agente [Amazon CloudWatch](#).

O Ganglia informa as métricas do Hadoop para cada instância. Os vários tipos de métricas são prefixados por categoria: sistema de arquivos distribuído (dfs.*), máquina virtual Java (jvm.*), (mapred.*) e chamadas de procedimento remoto MapReduce (rpc.*).

As métricas do Ganglia com base no YARN, como o Spark e o Hadoop, não estão disponíveis para as versões 4.4.0 e 4.5.0 do EMR. Use uma versão mais recente para usar essas métricas.

As métricas do Ganglia para o Spark geralmente têm prefixos para o ID de aplicativo do YARN e para o DAGScheduler do Spark. Portanto, os prefixos sigam este formato:

- DAGScheduler.*
- application_XXXXXXXXXX_XXXX.driver.*
- application_XXXXXXXXXX_XXXX.executor.*

Histórico de versões do Ganglia

Note

A última versão do Amazon EMR a incluir o Ganglia foi o Amazon EMR 6.15.0. Para monitorar seu cluster, versões superiores à 6.15.0 incluem o agente [Amazon CloudWatch](#).

A tabela a seguir informa a versão do Ganglia incluída em cada versão do Amazon EMR, junto com os componentes instalados com a aplicação. Para obter as versões dos componentes em cada versão, consulte a seção Versão de componentes da sua versão em [Versões de lançamento do Amazon EMR 7.x](#), [Versões de lançamento do Amazon EMR 6.x](#) ou [Versões de lançamento do Amazon EMR 5.x](#).

Informações sobre a versão do Ganglia

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o Ganglia
emr-5.36.2	3.7.2	emrfs, emr-goodies, ganglia-monitor, ganglia-metadata-collector, ganglia-web, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o Ganglia
		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, gangliamonitor, 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, gangliamonitor, 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

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, webserver

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, webserver

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, webserver

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, webserver

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, webserver

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, webserver

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, webserver

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, webserver

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, webserver

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, webserver
emr-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-resourcemanager, hadoop-yarn-timeline-server, webserver

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, webserver

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, webserver

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, webserver

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, webserver

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, webserver

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, webserver

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, webserver

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, webserver

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, webserver

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, webserver

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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

Rótulo de versão do Amazon EMR	Versão do Ganglia	Componentes instalados com o 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

O [Apache Hadoop](#) é uma estrutura de software Java de código aberto compatível com processamento de dados em massa por todo um cluster de instâncias. Ele também pode ser executado em uma única instância ou em milhares de instâncias. O Hadoop usa vários modelos de processamento, como MapReduce e Tez, para distribuir o processamento em várias instâncias e também usa um sistema de arquivos distribuído chamado HDFS para armazenar dados em várias instâncias. O Hadoop monitora a integridade das instâncias no cluster e pode fazer a recuperação de falha em um ou mais nós. Dessa maneira, o Hadoop proporciona o aumento do processamento e da capacidade de armazenamento, bem como a alta disponibilidade. Para obter mais informações, consulte a [documentação do Hadoop](#).

A tabela a seguir lista a versão do Hadoop incluída na versão mais recente da série Amazon EMR 7.x, junto com os componentes que a Amazon EMR instala com o Hadoop.

Para a versão dos componentes instalados com o Hadoop nesta versão, consulte Versões de componentes da [versão 7.2.0](#).

Informações sobre a versão do Hadoop para o emr-7.2.0

Gravadora EMR de lançamento da Amazon	Versão do Hadoop	Componentes instalados com o 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-https-server, hadoop-kms-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server

A tabela a seguir lista a versão do Hadoop incluída na versão mais recente da série Amazon EMR 6.x, junto com os componentes que a Amazon EMR instala com o Hadoop.

Para obter a versão dos componentes instalados com o Hadoop nessa versão, consulte [Release 6.15.0 Component Versions](#).

Informações de versão do Hadoop para o emr-6.15.0

Gravadora EMR de lançamento da Amazon	Versão do Hadoop	Componentes instalados com o 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

A tabela a seguir lista a versão do Hadoop incluída na versão mais recente da série Amazon EMR 5.x, junto com os componentes que a Amazon EMR instala com o Hadoop.

Para a versão dos componentes instalados com o Hadoop nesta versão, consulte Versões de componentes da [versão 5.36.2](#).

Informações sobre a versão do Hadoop para emr-5.36.2

Gravadora EMR de lançamento da Amazon	Versão do Hadoop	Componentes instalados com o 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

Gravadora EMR de lançamento da Amazon	Versão do Hadoop	Componentes instalados com o Hadoop
		nager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server

A partir do Amazon EMR 5.18.0, você pode usar o repositório de EMR artefatos da Amazon para criar seu código de trabalho com base nas versões exatas das bibliotecas e dependências que estão disponíveis em versões específicas da Amazon. Para obter mais informações, consulte [Verificar dependências usando o repositório de artefatos do Amazon EMR](#).

Tópicos

- [Configurar o Hadoop](#)
- [Criptografia transparente HDFS na Amazon EMR](#)
- [Criar ou executar uma aplicação Hadoop](#)
- [Leia objetos restaurados](#)
- [Ative o reconhecimento de acesso não uniforme à memória para YARN contêineres](#)
- [Histórico de versões do Hadoop](#)

Configurar o Hadoop

As seções a seguir fornecem as configurações padrão para daemons, tarefas e HDFS

Tópicos

- [Configuração da tarefa](#)
- [Definições de configuração de daemon do Hadoop](#)
- [HDFS configuração](#)

Configuração da tarefa

Você pode definir variáveis de configuração para ajustar o desempenho de seus MapReduce trabalhos. Esta seção fornece os valores padrão para configurações importantes. Os valores padrão variam com base no tipo de EC2 instância do nó usado no cluster. HBase está disponível ao usar a

EMR versão 4.6.0 e posterior da Amazon. Diferentes padrões são usados quando HBase é instalado. Esses valores são fornecidos juntamente com os padrões iniciais.

O Hadoop 2 usa dois parâmetros, `mapreduce.map.java.opts` e `mapreduce.reduce.java.opts`, para configurar a memória para mapear e reduzir JVMs, respectivamente. Eles substituem a opção de configuração única `mapreduce.map.java.opts` nas versões anteriores do Hadoop.

Da mesma forma, `mapred.job.jvm.num.tasks` substitui `mapred.job.reuse.jvm.num.tasks` no Hadoop 2.7.2 e versões posteriores. A Amazon EMR define esse valor como 20, independentemente do tipo de EC2 instância. Você pode substituir essa configuração usando a classificação de configuração `mapred-site`. Definir um valor de -1 indica que a JVM pode ser reutilizado para um número infinito de tarefas em um único trabalho, e um valor de 1 indica que uma nova JVM é gerada para cada tarefa.

Por exemplo, para definir o valor de `mapred.job.jvm.num.tasks` como -1, você pode criar um arquivo com o seguinte conteúdo:

```
[
  {
    "Classification": "mapred-site",
    "Properties": {
      "mapred.job.jvm.num.tasks": "-1"
    }
  }
]
```

Ao usar o `create-cluster` comando ou `modify-instance-groups` comando do AWS CLI, você pode então referenciar o arquivo JSON de configuração. No exemplo a seguir, o arquivo de configuração é salvo como `myConfig.json` e armazenado no Amazon S3.

Note

Os caracteres de continuação de linha do Linux (`\`) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (`^`).

```
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
```

Você pode alterar os valores padrão listados abaixo usando a classificação de `mapred-site` configuração da mesma forma e definir vários valores e várias classificações de configuração usando um único JSON arquivo. Para obter mais informações, consulte [Configurar aplicações](#).

Com a Amazon EMR versão 5.21.0 e posterior, você pode substituir as configurações do cluster e especificar classificações de configuração adicionais para cada grupo de instâncias em um cluster em execução. Você faz isso usando o EMR console da Amazon, o AWS Command Line Interface (AWS CLI) ou AWS SDK o. Para obter mais informações, consulte [Supplying a Configuration for an Instance Group in a Running Cluster](#).

Valores padrão para definições de configuração de tarefa

Instance Types (Tipos de instâncias)

- [Instâncias c1](#)
- [Instâncias c3](#)
- [Instâncias c4](#)
- [Instâncias c5](#)
- [Instâncias c5a](#)
- [Instâncias c5ad](#)
- [Instâncias c5d](#)
- [Instâncias c5n](#)
- [Instâncias c6a](#)
- [Instâncias c6g](#)
- [Instâncias c6gd](#)
- [Instâncias c6gn](#)
- [Instâncias c6i](#)
- [Instâncias c6id](#)
- [Instâncias c6in](#)

- [Instâncias c7a](#)
- [Instâncias c7g](#)
- [Instâncias c7gd](#)
- [Instâncias c7gn](#)
- [Instâncias C7i](#)
- [Instâncias d2](#)
- [Instâncias d3](#)
- [Instâncias d3en](#)
- [Instâncias g3](#)
- [Instâncias g3s](#)
- [Instâncias do g4dn](#)
- [Instâncias g5](#)
- [Instâncias h1](#)
- [Instâncias i2](#)
- [Instâncias i3](#)
- [Instâncias i3en](#)
- [Instâncias i4g](#)
- [Instâncias i4i](#)
- [Instâncias im4gn](#)
- [Instâncias is4gen](#)
- [Instâncias m1](#)
- [Instâncias m2](#)
- [Instâncias m3](#)
- [Instâncias m4](#)
- [Instâncias m5](#)
- [Instâncias m5a](#)
- [Instâncias m5ad](#)
- [Instâncias m5d](#)
- [Instâncias m5dn](#)

- [Instâncias m5n](#)
- [instâncias m5zn](#)
- [Instâncias m6a](#)
- [Instâncias m6g](#)
- [Instâncias m6gd](#)
- [Instâncias m6i](#)
- [Instâncias m6id](#)
- [Instâncias m6idn](#)
- [Instâncias m6in](#)
- [Instâncias M7a](#)
- [Instâncias m7g](#)
- [Instâncias m7gd](#)
- [Instâncias M7i](#)
- [Instâncias M7i-flex](#)
- [Instâncias p2](#)
- [Instâncias p3](#)
- [Instâncias p5](#)
- [Instâncias r3](#)
- [Instâncias r4](#)
- [Instâncias r5](#)
- [Instâncias r5a](#)
- [Instâncias r5ad](#)
- [Instâncias r5b](#)
- [Instâncias r5d](#)
- [Instâncias r5dn](#)
- [Instâncias r5n](#)
- [Instâncias r6a](#)
- [Instâncias r6g](#)
- [Instâncias r6gd](#)

- [Instâncias r6i](#)
- [Instâncias r6iD](#)
- [instâncias r6idn](#)
- [Instâncias r6in](#)
- [Instâncias R7a](#)
- [Instâncias r7g](#)
- [Instâncias r7gd](#)
- [Instâncias r7i](#)
- [Instâncias R7iz](#)
- [Instâncias x1](#)
- [Instâncias x1e](#)
- [Instâncias x2gd](#)
- [instâncias x2idn](#)
- [instâncias x2iedn](#)
- [Instâncias z1d](#)

Instâncias c1

c1.medium

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<code>yarn.scheduler.maximum-allocation-mb</code>	512	512
<code>yarn.nodemanager.resource-memory-mb</code>	1024	512

c1.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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>	1024	1024
<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

Instâncias c3

c3.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<code>mapreduce.map.java.opts</code>	<code>-Xmx1126m</code>	<code>-Xmx1126m</code>

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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>	2.880	2.880
<code>yarn.app.mapreduce.am.resource.mb</code>	2.880	2.880
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	11520	5760

Opção de configuração	Valor padrão	Com HBase instalado
<code>yarn.nodemanager.resource.memory-mb</code>	11520	5760

c3.4xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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>	2.880	2.880
<code>yarn.app.mapreduce.am.resource.mb</code>	2.880	2.880
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias c4

c4.large

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<code>yarn.nodemanager.resource.memory-mb</code>	1792	896

c4.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

c4.2xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<code>mapreduce.reduce.memory.mb</code>	2.880	2.880
<code>yarn.app.mapreduce.am.resource.mb</code>	2.880	2.880
<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

c4.4xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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>	2.880	2.880
<code>yarn.app.mapreduce.am.resource.mb</code>	2.880	2.880
<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.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<code>mapreduce.map.java.opts</code>	<code>-Xmx1183m</code>	<code>-Xmx1183m</code>
<code>mapreduce.java.opts</code>	<code>-Xmx2366m</code>	<code>-Xmx2366m</code>
<code>mapreduce.map.memory.mb</code>	1479	1479
<code>mapreduce.reduce.memory.mb</code>	2958	2958
<code>yarn.app.mapreduce.am.resource.mb</code>	2958	2958
<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

Instâncias c5

c5.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias c5a

c5a.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias c5ad

c5ad.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	5632	2816
yarn.nodemanager.resource.memory-mb	5632	2816

c5ad.2xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Instâncias c5d

c5d.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24576	12288
yarn.nodemanager.resource.memory-mb	24576	12288

c5d.9xlarge

Opção de configuração	Valor padrão	Com HBase instalado
mapreduce.map.java.opts	-Xmx1456m	-Xmx1456m
mapreduce.java.opts	-Xmx2912m	-Xmx2912m
mapreduce.map.memory.mb	1820	1820
mapreduce.reduce.memory.mb	3640	3640
yarn.app.mapreduce.am.resource.mb	3640	3640

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
mapreduce.map.java.opts	-Xmx1547m	-Xmx1547m
mapreduce.java.opts	-Xmx3094m	-Xmx3094m
mapreduce.map.memory.mb	1934	1934
mapreduce.reduce.memory.mb	3868	3868
yarn.app.mapreduce.am.resource.mb	3868	3868
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	139264	30960
yarn.nodemanager.resource.memory-mb	139264	30960

c5d.24xlarge

Opção de configuração	Valor padrão	Com HBase instalado
mapreduce.map.java.opts	-Xmx1570m	-Xmx1570m
mapreduce.java.opts	-Xmx3140m	-Xmx3140m
mapreduce.map.memory.mb	1963	1963
mapreduce.reduce.memory.mb	3926	3926
yarn.app.mapreduce.am.resource.mb	3926	3926

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias c5n

c5n.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
mapreduce.map.java.opts	-Xmx1613m	-Xmx1613m
mapreduce.java.opts	-Xmx3226m	-Xmx3226m
mapreduce.map.memory.mb	2016	2016
mapreduce.reduce.memory.mb	4032	4032
yarn.app.mapreduce.am.resource.mb	4032	4032
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	16128	8064
yarn.nodemanager.resource.memory-mb	16128	8064

c5n.4xlarge

Opção de configuração	Valor padrão	Com HBase instalado
mapreduce.map.java.opts	-Xmx1741m	-Xmx1741m
mapreduce.java.opts	-Xmx3482m	-Xmx3482m
mapreduce.map.memory.mb	2176	2176
mapreduce.reduce.memory.mb	4352	4352
yarn.app.mapreduce.am.resource.mb	4352	4352

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Instâncias c6a

c6a.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

c6a.16xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

c6a.24xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Instâncias c6g

c6g.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

c6g.8xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Instâncias c6gd

c6gd.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Instâncias c6gn

c6gn.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Instâncias c6i

c6i.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

c6i.8xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias c6id

c6id.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias c6in

c6in.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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	11712	5856
yarn.nodemanager.resource.memory-mb	11712	5856

c6in.4xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Instâncias c7a

c7a.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

c7a.8xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

c7a.16xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Instâncias c7g

c7g.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Instâncias c7gd

c7gd.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

c7gd.8xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Instâncias c7gn

c7gn.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

c7gn.8xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Instâncias C7i

c7i.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

c7i.8xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias d2

d2.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias d3

d3.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias d3en

d3en.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

d3en.4xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	2980
<code>yarn.nodemanager.resource.memory-mb</code>	179200	2980

Instâncias g3

g3.4xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias g3s

g3s.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias do g4dn

g4dn.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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>	3968	3968
<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>	253952	31744
<code>yarn.nodemanager.resource.memory-mb</code>	253952	31744

Instâncias g5

g5.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

g5.12xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	179200	2980
<code>yarn.nodemanager.resource-memory-mb</code>	179200	2980

g5.16xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

g5.48xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias h1

h1.2xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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>	3968	3968
<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>	253952	31744
<code>yarn.nodemanager.resource-memory-mb</code>	253952	31744

Instâncias i2

i2.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

i2.8xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias i3

i3.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias i3en

i3en.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias i4g

i4g.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias i4i

i4i.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
mapreduce.map.java.opts	-Xmx6195m	-Xmx6195m
mapreduce.java.opts	-Xmx12390m	-Xmx12390m
mapreduce.map.memory.mb	774	774
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

Instâncias im4gn

im4gn.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

im4gn.8xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias is4gen

is4gen.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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	8.786,06
<code>yarn.nodemanager.resource.memory-mb</code>	17572,12	8.786,06

is4gen.2xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
mapreduce.map.java.opts	-Xmx4480m	-Xmx4480m
mapreduce.java.opts	-Xmx8960m	-Xmx8960m
mapreduce.map.memory.mb	5600	5600
mapreduce.reduce.memory.mb	11200	11200
yarn.app.mapreduce.am.resource.mb	11200	11200
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	179200	22400
yarn.nodemanager.resource.memory-mb	179200	22400

Instâncias m1

m1.small

Opção de configuração	Valor padrão	Com HBase instalado
mapreduce.map.java.opts	-Xmx288m	-Xmx288m
mapreduce.java.opts	-Xmx288m	-Xmx288m
mapreduce.map.memory.mb	512	512
mapreduce.reduce.memory.mb	512	512
yarn.app.mapreduce.am.resource.mb	512	512

Opção de configuração	Valor padrão	Com HBase instalado
<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>	1024	512

m1.medium

Opção de configuração	Valor padrão	Com HBase instalado
<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>	1024	1024
<code>yarn.app.mapreduce.am.resource.mb</code>	1024	1024
<code>yarn.scheduler.minimum-allocation-mb</code>	256	256
<code>yarn.scheduler.maximum-allocation-mb</code>	2048	1024
<code>yarn.nodemanager.resource.memory-mb</code>	2048	1024

m1.large

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias m2

m2.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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>	1024	1024
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias m3

m3.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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>	2.880	2.880
<code>yarn.app.mapreduce.am.resource.mb</code>	2.880	2.880
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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>	2.880	2.880
<code>yarn.app.mapreduce.am.resource.mb</code>	2.880	2.880
<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

Instâncias m4

m4.large

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	24576	12288
yarn.nodemanager.resource.memory-mb	24576	12288

m4.4xlarge

Opção de configuração	Valor padrão	Com HBase instalado
mapreduce.map.java.opts	-Xmx1434m	-Xmx1434m
mapreduce.java.opts	-Xmx2868m	-Xmx2868m
mapreduce.map.memory.mb	1792	1792
mapreduce.reduce.memory.mb	3584	3584
yarn.app.mapreduce.am.resource.mb	3584	3584

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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>	1984	1984
<code>mapreduce.reduce.memory.mb</code>	3968	3968
<code>yarn.app.mapreduce.am.resource.mb</code>	3968	3968
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	253952	31744
<code>yarn.nodemanager.resource.memory-mb</code>	253952	31744

Instâncias m5

m5.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
mapreduce.map.java.opts	-Xmx3174m	-Xmx3174m
mapreduce.java.opts	-Xmx6348m	-Xmx6348m
mapreduce.map.memory.mb	3968	3968
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	253952	31744
yarn.nodemanager.resource.memory-mb	253952	31744

m5.24xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias m5a

m5a.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	188416	31416
yarn.nodemanager.resource.memory-mb	188416	31416

m5a.16xlarge

Opção de configuração	Valor padrão	Com HBase instalado
mapreduce.map.java.opts	-Xmx3174m	-Xmx3174m
mapreduce.java.opts	-Xmx6348m	-Xmx6348m
mapreduce.map.memory.mb	3968	3968
mapreduce.reduce.memory.mb	7936	7936
yarn.app.mapreduce.am.resource.mb	7936	7936

Opção de configuração	Valor padrão	Com HBase instalado
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	253952	31744
<code>yarn.nodemanager.resource.memory-mb</code>	253952	31744

m5a.24xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias m5ad

m5ad.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Instâncias m5d

m5d.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
mapreduce.map.java.opts	-Xmx3174m	-Xmx3174m
mapreduce.java.opts	-Xmx6348m	-Xmx6348m
mapreduce.map.memory.mb	3968	3968
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	253952	31744
yarn.nodemanager.resource.memory-mb	253952	31744

m5d.24xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias m5dn

m5dn.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

m5dn.4xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias m5n

m5n.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

instâncias m5zn

m5zn.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
mapreduce.map.java.opts	-Xmx2304m	-Xmx2304m
mapreduce.java.opts	-Xmx4608m	-Xmx4608m
mapreduce.map.memory.mb	2.880	2.880
mapreduce.reduce.memory.mb	5760	5760
yarn.app.mapreduce.am.resource.mb	5760	5760

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias m6a

m6a.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias m6g

m6g.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias m6gd

m6gd.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias m6i

m6i.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

m6i.12xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias m6id

m6id.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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	2980
yarn.nodemanager.resource.memory-mb	179200	2980

m6id.16xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Instâncias m6idn

m6idn.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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	2980
<code>yarn.nodemanager.resource.memory-mb</code>	179200	2980

m6idn.16xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias m6in

m6in.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias M7a

m7a.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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	2980
yarn.nodemanager.resource.memory-mb	179200	2980

m7a.16xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias m7g

m7g.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

m7g.4xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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	2980
yarn.nodemanager.resource.memory-mb	179200	2980

m7g.16xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias m7gd

m7gd.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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	2980
yarn.nodemanager.resource.memory-mb	179200	2980

m7gd.16xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias M7i

m7i.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

m7i.4xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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	181248	30208
yarn.nodemanager.resource.memory-mb	181248	30208

m7i.16xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Instâncias M7i-flex

m7i-flex.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias p2

p2.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias p3

p3.2xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

p3.16xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias p5

p5.48xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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>	1990-656	20736
<code>yarn.nodemanager.resource.memory-mb</code>	1990-656	20736

Instâncias r3

r3.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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.resou rce.mb</code>	5856	5856
<code>yarn.scheduler.minimum-allo cation-mb</code>	32	32
<code>yarn.scheduler.maximum-allo cation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.m emory-mb</code>	23424	11712

r3.2xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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.resou rce.mb</code>	6784	6784

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

r3.8xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias r4

r4.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias r5

r5.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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>	253952	31744
<code>yarn.nodemanager.resource.memory-mb</code>	253952	31744

r5.12xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias r5a

r5a.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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	122880	30720
yarn.nodemanager.resource.memory-mb	122880	30720

r5a.8xlarge

Opção de configuração	Valor padrão	Com HBase instalado
mapreduce.map.java.opts	-Xmx6349m	-Xmx6349m
mapreduce.java.opts	-Xmx12698m	-Xmx12698m
mapreduce.map.memory.mb	7936	7936
mapreduce.reduce.memory.mb	15872	15872
yarn.app.mapreduce.am.resource.mb	15872	15872

Opção de configuração	Valor padrão	Com HBase instalado
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	253952	31744
<code>yarn.nodemanager.resource.memory-mb</code>	253952	31744

r5a.12xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias r5ad

r5ad.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias r5b

r5b.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

r5b.12xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Instâncias r5d

r5d.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	253952	31744
<code>yarn.nodemanager.resource.memory-mb</code>	253952	31744

r5d.12xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

r5d.24xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias r5dn

r5dn.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

r5dn.16xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias r5n

r5n.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Instâncias r6a

r6a.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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>	774	774
<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

Opção de configuração	Valor padrão	Com HBase instalado
mapreduce.map.java.opts	-Xmx6212m	-Xmx6212m
mapreduce.java.opts	-Xmx12424m	-Xmx12424m
mapreduce.map.memory.mb	7765	7765
mapreduce.reduce.memory.mb	1530	1530
yarn.app.mapreduce.am.resource.mb	1530	1530
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	1490944	31124
yarn.nodemanager.resource.memory-mb	1490944	31124

Instâncias r6g

r6g.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

r6g.8xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Instâncias r6gd

r6gd.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

r6gd.8xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Instâncias r6i

r6i.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

r6i.8xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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	3014
<code>yarn.nodemanager.resource.memory-mb</code>	964608	3014

Instâncias r6iD

r6id.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
mapreduce.map.java.opts	-Xmx6195m	-Xmx6195m
mapreduce.java.opts	-Xmx12390m	-Xmx12390m
mapreduce.map.memory.mb	774	774
mapreduce.reduce.memory.mb	1548	1548
yarn.app.mapreduce.am.resource.mb	1548	1548

Opção de configuração	Valor padrão	Com HBase instalado
<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

instâncias r6idn

r6idn.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
mapreduce.map.java.opts	-Xmx6195m	-Xmx6195m
mapreduce.java.opts	-Xmx12390m	-Xmx12390m
mapreduce.map.memory.mb	774	774
mapreduce.reduce.memory.mb	1548	1548
yarn.app.mapreduce.am.resource.mb	1548	1548
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	991232	30976
yarn.nodemanager.resource.memory-mb	991232	30976

Instâncias r6in

r6in.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

r6in.24xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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>	774	774
<code>mapreduce.reduce.memory.mb</code>	1548	1548
<code>yarn.app.mapreduce.am.resource.mb</code>	1548	1548
<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

Instâncias R7a

r7a.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
mapreduce.map.java.opts	-Xmx6195m	-Xmx6195m
mapreduce.java.opts	-Xmx12390m	-Xmx12390m
mapreduce.map.memory.mb	774	774
mapreduce.reduce.memory.mb	15488	15488
yarn.app.mapreduce.am.resource.mb	15488	15488

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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>	1530	1530
<code>yarn.app.mapreduce.am.resource.mb</code>	1530	1530
<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

Instâncias r7g

r7g.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias r7gd

r7gd.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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.resou rce.mb</code>	11712	11712
<code>yarn.scheduler.minimum-allo cation-mb</code>	32	32
<code>yarn.scheduler.maximum-allo cation-mb</code>	23424	11712
<code>yarn.nodemanager.resource.m emory-mb</code>	23424	11712

r7gd.2xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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.resou rce.mb</code>	13568	13568

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias r7i

r7i.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

r7i.12xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

r7i.48xlarge

Opção de configuração	Valor padrão	Com HBase instalado
mapreduce.map.java.opts	-Xmx6212m	-Xmx6212m
mapreduce.java.opts	-Xmx12424m	-Xmx12424m
mapreduce.map.memory.mb	7765	7765
mapreduce.reduce.memory.mb	1530	1530
yarn.app.mapreduce.am.resource.mb	1530	1530

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias R7iz

r7iz.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

r7iz.4xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

r7iz.16xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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	3014
<code>yarn.nodemanager.resource.memory-mb</code>	964608	3014

Instâncias x1

x1.16xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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>	3014	3014
<code>yarn.app.mapreduce.am.resource.mb</code>	3014	3014
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	964608	3014
<code>yarn.nodemanager.resource.memory-mb</code>	964608	3014

x1.32xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	1937-408	30272
<code>yarn.nodemanager.resource.memory-mb</code>	1937-408	30272

Instâncias x1e

x1e.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
mapreduce.map.java.opts	-Xmx23501m	-Xmx23501m
mapreduce.java.opts	-Xmx47002m	-Xmx47002m
mapreduce.map.memory.mb	29376	29376
mapreduce.reduce.memory.mb	58752	58752
yarn.app.mapreduce.am.resource.mb	58752	58752
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	235008	0
yarn.nodemanager.resource.memory-mb	235008	0

x1e.4xlarge

Opção de configuração	Valor padrão	Com HBase instalado
mapreduce.map.java.opts	-Xmx23910m	-Xmx23910m
mapreduce.java.opts	-Xmx47820m	-Xmx47820m
mapreduce.map.memory.mb	2988	2988
mapreduce.reduce.memory.mb	59776	59776
yarn.app.mapreduce.am.resource.mb	59776	59776

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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>	3014	3014
<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

Opção de configuração	Valor padrão	Com HBase instalado
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	1937-408	0
yarn.nodemanager.resource.memory-mb	1937-408	0

x1e.32xlarge

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias x2gd

x2gd.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
mapreduce.map.java.opts	-Xmx11674m	-Xmx11674m
mapreduce.java.opts	-Xmx23348m	-Xmx23348m
mapreduce.map.memory.mb	14592	14592
mapreduce.reduce.memory.mb	29184	29184
yarn.app.mapreduce.am.resource.mb	29184	29184
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	116736	29184
yarn.nodemanager.resource.memory-mb	116736	29184

x2gd.4xlarge

Opção de configuração	Valor padrão	Com HBase instalado
mapreduce.map.java.opts	-Xmx12083m	-Xmx12083m
mapreduce.java.opts	-Xmx24166m	-Xmx24166m
mapreduce.map.memory.mb	15104	15104
mapreduce.reduce.memory.mb	30208	30208
yarn.app.mapreduce.am.resource.mb	30208	30208

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
mapreduce.map.java.opts	-Xmx12390m	-Xmx12390m
mapreduce.java.opts	-Xmx24780m	-Xmx24780m
mapreduce.map.memory.mb	1548	1548
mapreduce.reduce.memory.mb	30976	30976
yarn.app.mapreduce.am.resource.mb	30976	30976

Opção de configuração	Valor padrão	Com HBase instalado
<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

instâncias x2idn

x2idn.16xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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>	1548	1548
<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

Opção de configuração	Valor padrão	Com HBase instalado
mapreduce.map.java.opts	-Xmx12425m	-Xmx12425m
mapreduce.java.opts	-Xmx24850m	-Xmx24850m
mapreduce.map.memory.mb	15531	15531
mapreduce.reduce.memory.mb	31062	31062
yarn.app.mapreduce.am.resource.mb	31062	31062
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	1490944	31030
yarn.nodemanager.resource.memory-mb	1490944	31030

x2idn.32xlarge

Opção de configuração	Valor padrão	Com HBase instalado
mapreduce.map.java.opts	-Xmx12442m	-Xmx12442m
mapreduce.java.opts	-Xmx24884m	-Xmx24884m
mapreduce.map.memory.mb	1552	1552
mapreduce.reduce.memory.mb	31104	31104
yarn.app.mapreduce.am.resource.mb	31104	31104

Opção de configuração	Valor padrão	Com HBase instalado
<code>yarn.scheduler.minimum-allocation-mb</code>	32	32
<code>yarn.scheduler.maximum-allocation-mb</code>	1990-656	31104
<code>yarn.nodemanager.resource.memory-mb</code>	1990-656	31104

instâncias x2iedn

x2iedn.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
mapreduce.map.java.opts	-Xmx24883m	-Xmx24883m
mapreduce.java.opts	-Xmx49766m	-Xmx49766m
mapreduce.map.memory.mb	31104	31104
mapreduce.reduce.memory.mb	62208	62208
yarn.app.mapreduce.am.resource.mb	62208	62208
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	1990-656	0
yarn.nodemanager.resource.memory-mb	1990-656	0

x2iedn.24xlarge

Opção de configuração	Valor padrão	Com HBase instalado
mapreduce.map.java.opts	-Xmx24918m	-Xmx24918m
mapreduce.java.opts	-Xmx49836m	-Xmx49836m
mapreduce.map.memory.mb	31147	31147
mapreduce.reduce.memory.mb	6294	6294
yarn.app.mapreduce.am.resource.mb	6294	6294

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Instâncias z1d

z1d.xlarge

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
<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

Opção de configuração	Valor padrão	Com HBase instalado
mapreduce.map.java.opts	-Xmx6281m	-Xmx6281m
mapreduce.java.opts	-Xmx12562m	-Xmx12562m
mapreduce.map.memory.mb	7851	7851
mapreduce.reduce.memory.mb	15702	15702
yarn.app.mapreduce.am.resource.mb	15702	15702
yarn.scheduler.minimum-allocation-mb	32	32
yarn.scheduler.maximum-allocation-mb	188416	31396
yarn.nodemanager.resource.memory-mb	188416	31396

z1d.12xlarge

Opção de configuração	Valor padrão	Com HBase instalado
mapreduce.map.java.opts	-Xmx6417m	-Xmx6417m
mapreduce.java.opts	-Xmx12834m	-Xmx12834m
mapreduce.map.memory.mb	8021	8021
mapreduce.reduce.memory.mb	16042	16042
yarn.app.mapreduce.am.resource.mb	16042	16042

Opção de configuração	Valor padrão	Com HBase instalado
<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

Definições de configuração de daemon do Hadoop

As configurações de daemon do Hadoop são diferentes, dependendo do tipo de instância do EC2 que um nó de cluster usa. As tabelas a seguir listam as definições de configuração padrão para cada tipo de instância do EC2.

Para personalizar essas configurações, use a classificação de configuração `hadoop-env`. Para obter mais informações, consulte [Configurar aplicações](#).

Instance Types (Tipos de instâncias)

- [Instâncias c1](#)
- [Instâncias c3](#)
- [Instâncias c4](#)
- [Instâncias c5](#)
- [Instâncias c5a](#)
- [Instâncias c5ad](#)
- [Instâncias c5d](#)
- [Instâncias c5n](#)
- [Instâncias c6a](#)
- [Instâncias c6g](#)
- [Instâncias c6gd](#)
- [Instâncias c6gn](#)
- [Instâncias c6i](#)

- [Instâncias c6id](#)
- [Instâncias c6in](#)
- [Instâncias c7a](#)
- [Instâncias c7g](#)
- [Instâncias c7gd](#)
- [Instâncias c7gn](#)
- [Instâncias C7i](#)
- [Instâncias d2](#)
- [Instâncias d3](#)
- [Instâncias d3en](#)
- [Instâncias g3](#)
- [Instâncias g3s](#)
- [Instâncias do g4dn](#)
- [Instâncias g5](#)
- [Instâncias h1](#)
- [Instâncias i2](#)
- [Instâncias i3](#)
- [Instâncias i3en](#)
- [Instâncias i4g](#)
- [Instâncias i4i](#)
- [Instâncias im4gn](#)
- [Instâncias is4gen](#)
- [Instâncias m1](#)
- [Instâncias m2](#)
- [Instâncias m3](#)
- [Instâncias m4](#)
- [Instâncias m5](#)
- [Instâncias m5a](#)
- [Instâncias m5ad](#)

- [Instâncias m5d](#)
- [Instâncias m5dn](#)
- [Instâncias m5n](#)
- [instâncias m5zn](#)
- [Instâncias m6a](#)
- [Instâncias m6g](#)
- [Instâncias m6gd](#)
- [Instâncias m6i](#)
- [Instâncias m6id](#)
- [Instâncias m6idn](#)
- [Instâncias m6in](#)
- [Instâncias M7a](#)
- [Instâncias m7g](#)
- [Instâncias m7gd](#)
- [Instâncias m7i](#)
- [Instâncias m7i-flex](#)
- [Instâncias p2](#)
- [Instâncias p3](#)
- [Instâncias p5](#)
- [Instâncias r3](#)
- [Instâncias r4](#)
- [Instâncias r5](#)
- [Instâncias r5a](#)
- [Instâncias r5ad](#)
- [Instâncias r5b](#)
- [Instâncias r5d](#)
- [Instâncias r5dn](#)
- [Instâncias r5n](#)
- [Instâncias r6a](#)
- [Instâncias r6g](#)

- [Instâncias r6gd](#)
- [Instâncias r6i](#)
- [Instâncias r6iD](#)
- [instâncias r6idn](#)
- [Instâncias r6in](#)
- [Instâncias R7a](#)
- [Instâncias r7g](#)
- [Instâncias r7gd](#)
- [Instâncias r7i](#)
- [Instâncias R7iz](#)
- [Instâncias x1](#)
- [Instâncias x1e](#)
- [Instâncias x2gd](#)
- [instâncias x2idn](#)
- [instâncias x2iedn](#)
- [Instâncias z1d](#)

Instâncias c1

c1.medium

Parâmetro	Valor
YARN_RESOURCEMANAGER_HEAPSIZE	192
YARN_PROXYSERVER_HEAPSIZE	96
YARN_NODEMANAGER_HEAPSIZE	128
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	128
HADOOP_NAMENODE_HEAPSIZE	192

Parâmetro	Valor
HADOOP_DATANODE_HEAPSIZE	96

c1.xlarge

Parâmetro	Valor
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

Instâncias c3

c3.xlarge

Parâmetro	Valor
YARN_RESOURCEMANAGER_HEAPSIZE	2124
YARN_PROXYSERVER_HEAPSIZE	2124
YARN_NODEMANAGER_HEAPSIZE	2048
HADOOP_JOB_HISTORYSERVER_HEAPSIZE	2124
HADOOP_NAMENODE_HEAPSIZE	972

Parâmetro	Valor
HADOOP_DATANODE_HEAPSIZE	588

c3.2xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

c3.8xlarge

Parâmetro	Valor
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

Instâncias c4

c4.large

Parâmetro	Valor
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

Parâmetro	Valor
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

c4.2xlarge

Parâmetro	Valor
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

c4.4xlarge

Parâmetro	Valor
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

c4.8xlarge

Parâmetro	Valor
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

Instâncias c5

c5.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias c5a

c5a.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias c5ad

c5ad.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias c5d

c5d.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias c5n

c5n.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias c6a

c6a.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias c6g

c6g.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias c6gd

c6gd.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias c6gn

c6gn.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias c6i

c6i.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias c6id

c6id.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias c6in

c6in.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias c7a

c7a.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias c7g

c7g.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias c7gd

c7gd.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias c7gn

c7gn.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias C7i

c7i.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias d2

d2.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias d3

d3.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias d3en

d3en.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias g3

g3.4xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias g3s

g3s.xlarge

Parâmetro	Valor
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

Instâncias do g4dn

g4dn.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias g5

g5.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias h1

h1.2xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias i2

i2.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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.xlarge

Parâmetro	Valor
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

Instâncias i3

i3.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias i3en

i3en.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias i4g

i4g.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias i4i

i4i.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias im4gn

im4gn.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias is4gen

is4gen.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias m1

m1.small

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias m2

m2.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias m3

m3.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias m4

m4.large

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias m5

m5.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias m5a

m5a.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias m5ad

m5ad.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias m5d

m5d.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias m5dn

m5dn.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias m5n

m5n.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

instâncias m5zn

m5zn.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias m6a

m6a.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias m6g

m6g.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias m6gd

m6gd.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias m6i

m6i.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias m6id

m6id.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias m6idn

m6idn.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias m6in

m6in.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias M7a

m7a.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias m7g

m7g.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias m7gd

m7gd.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias m7i

m7i.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias m7i-flex

m7i-flex.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias p2

p2.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias p3

p3.2xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias p5

p5.48xlarge

Parâmetro	Valor
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

Instâncias r3

r3.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias r4

r4.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias r5

r5.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias r5a

r5a.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias r5ad

r5ad.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias r5b

r5b.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias r5d

r5d.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias r5dn

r5dn.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias r5n

r5n.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias r6a

r6a.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias r6g

r6g.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias r6gd

r6gd.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias r6i

r6i.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias r6iD

r6id.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

instâncias r6idn

r6idn.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias r6in

r6in.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias R7a

r7a.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias r7g

r7g.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias r7gd

r7gd.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias r7i

r7i.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias R7iz

r7iz.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias x1

x1.16xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias x1e

x1e.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias x2gd

x2gd.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

instâncias x2idn

x2idn.16xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

instâncias x2iedn

x2iedn.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Instâncias z1d

z1d.xlarge

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

Parâmetro	Valor
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

HDFSconfiguração

A tabela a seguir descreve os parâmetros padrão do Hadoop Distributed File System (HDFS) e suas configurações. Você pode alterar esses valores usando a classificação de configuração `hdfs-site`. Para obter mais informações, consulte [Configurar aplicações](#).

Warning

1. `dfs.replication` Definir como 1 em clusters com menos de quatro nós pode levar à perda de HDFS dados se um único nó ficar inativo. Se seu cluster tiver HDFS armazenamento, recomendamos que você configure o cluster com pelo menos quatro nós principais para cargas de trabalho de produção, a fim de evitar perda de dados.
2. A Amazon não EMR permitirá que os clusters escalem os nós principais abaixo de `dfs.replication`. Por exemplo, se `dfs.replication = 2`, o número mínimo de nós central será 2.
3. Ao usar o Ajuste de Escala Gerenciado, o ajuste de escala automático ou optar por redimensionar manualmente o cluster, é recomendável definir `dfs.replication` como 2 ou mais.

Parâmetro	Definição	Valor padrão
<code>dfs.block.size</code>	O tamanho dos HDFS blocos. Ao operar com dados armazenados em HDFS, o tamanho da divisão geralmente é o tamanho de um HDFS bloco. Números maiores fornecem menos granularidade de tarefas, mas também impõem menos pressão no cluster NameNode.	134217728 (128 MB)
<code>dfs.replication</code>	O número de cópias de cada bloco a ser armazenado para durabilidade. A Amazon EMR define esse valor com base no número de nós principais com os quais o cluster é provisionado. Ajuste o valor para atender a suas necessidades. Para substituir o valor padrão, use a classificação <code>hdfs-site</code> .	<p>1 para clusters provisionados com menos de quatro nós centrais</p> <p>2 para clusters provisionados com menos de dez nós centrais</p> <p>3 para todos os outros clusters</p>

Criptografia transparente HDFS na Amazon EMR

A criptografia transparente é implementada por meio do uso de zonas de HDFS criptografia, que são HDFS caminhos definidos por você. Cada zona de criptografia tem sua própria chave, que é armazenada no servidor de chaves especificado usando a classificação da configuração `hdfs-site`.

A partir da EMR versão 4.8.0 da Amazon, você pode usar as configurações de EMR segurança da Amazon para definir as configurações de criptografia de dados para clusters com mais facilidade. As configurações de segurança oferecem configurações para permitir a segurança de dados em trânsito e dados em repouso nos volumes de armazenamento e dados do Amazon Elastic Block Store (AmazonEBS) no Amazon EMRFS S3. Para obter mais informações, consulte [Criptografar dados em trânsito e em repouso](#) no Guia EMR de gerenciamento da Amazon.

A Amazon EMR usa o Hadoop KMS por padrão; no entanto, você pode usar outro KMS que implemente a operação. KeyProvider API Cada arquivo em uma zona de HDFS criptografia tem sua própria chave de criptografia de dados exclusiva, que é criptografada pela chave da zona de criptografia. HDFSos dados são criptografados end-to-end (em repouso e em trânsito) quando são gravados em uma zona de criptografia porque as atividades de criptografia e descriptografia ocorrem somente no cliente.

Você não pode mover arquivos entre zonas de criptografias ou de uma zona de criptografia para caminhos não criptografados.

O HDFS cliente NameNode e interagem com o Hadoop KMS (ou uma alternativa KMS que você configurou) por meio da KeyProvider API operação. O KMS é responsável por armazenar as chaves de criptografia no repositório de chaves de apoio. Além disso, a Amazon EMR inclui a política de força JCE ilimitada, para que você possa criar chaves no tamanho desejado.

Para obter mais informações, consulte [Criptografia transparente HDFS na](#) documentação do Hadoop.

Note

Na AmazonEMR, KMS o over não HTTPS está habilitado por padrão com o HadoopKMS. Para obter mais informações sobre como habilitar o KMS overHTTPS, consulte a documentação do [Hadoop KMS](#).

Configurando a criptografia HDFS transparente

Você pode configurar a criptografia transparente na Amazon EMR criando chaves e adicionando zonas de criptografia. Você pode fazer isso de várias maneiras:

- Usando a API operação EMR de configuração da Amazon ao criar um cluster
- Usando uma JAR etapa do Hadoop com command-runner.jar
- Fazendo login no nó principal do cluster Hadoop e usando os clientes de linha de comando `hadoop key` e `hdfs crypto`
- Usando o REST APIs para Hadoop KMS e HDFS

Para obter mais informações sobre o REST APIs, consulte a respectiva documentação do Hadoop e. KMS HDFS

Para criar zonas de criptografia e suas chaves na criação do cluster usando o CLI

A `hdfs-encryption-zones` classificação na API operação de configuração permite que você especifique um nome de chave e uma zona de criptografia ao criar um cluster. EMRA Amazon cria essa chave no Hadoop KMS em seu cluster e configura a zona de criptografia.

- Crie um cluster com o comando a seguir.

```
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

Os caracteres de continuação de linha do Linux (`\`) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (`^`).

myConfig.json:

```
[
  {
    "Classification": "hdfs-encryption-zones",
    "Properties": {
      "/myHDFSPath1": "path1_key",
      "/myHDFSPath2": "path2_key"
    }
  }
]
```

Para criar zonas de criptografia e suas chaves manualmente no nó principal

1. Inicie seu cluster usando uma EMR versão da Amazon maior que 4.1.0.
2. Conecte-se ao nó principal do cluster com SSH.
3. Crie uma chave no 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

O Hadoop KMS exige que seus nomes de chave estejam em minúsculas. Se você usar uma chave que tenha caracteres em maiúsculas, seu cluster falhará durante a execução.

4. Crie o caminho da zona de criptografia emHDFS.

```
$ hadoop fs -mkdir /myHDFSPath2
```

5. Transforme o HDFS caminho em uma zona de criptografia usando a chave que você criou.

```
$ hdfs crypto -createZone -keyName path2_key -path /myHDFSPath2
Added encryption zone /myHDFSPath2
```

Para criar zonas de criptografia e suas chaves manualmente usando o AWS CLI

- Adicione etapas para criar as KMS chaves e as zonas de criptografia manualmente com o comando a seguir.

```
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

Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

Considerações sobre criptografia HDFS transparente

Uma prática recomendada é criar uma zona de criptografia para cada aplicativo, na qual eles podem gravar arquivos. Além disso, você pode criptografar tudo HDFS usando a `hdfs-encryption-zones` classificação na configuração API e especificar o caminho raiz (/) como a zona de criptografia.

Servidor de gerenciamento de chaves do Hadoop

O [Hadoop KMS](#) é um servidor de gerenciamento de chaves que fornece a capacidade de implementar serviços criptográficos para clusters do Hadoop e pode servir como provedor de chave para [Criptografia transparente HDFS na Amazon EMR](#). O Hadoop KMS na Amazon EMR é instalado e habilitado por padrão quando você seleciona o aplicativo Hadoop ao iniciar um cluster. EMR O Hadoop KMS não armazena as chaves em si, exceto no caso de armazenamento temporário em cache. O Hadoop KMS atua como um proxy entre o provedor da chave e o administrador do cliente para um armazenamento de chaves de apoio — ele não é um armazenamento de chaves. O repositório de chaves padrão criado para o Hadoop KMS é a Java Cryptography Extension (JCE) KeyStore JCEKS. A política de força JCE ilimitada também está incluída, para que você possa criar chaves com o tamanho desejado. O Hadoop KMS também oferece suporte a uma variedade de dispositivos ACLs que controlam o acesso às chaves e às operações principais, independentemente de outros aplicativos clientes, como HDFS. O tamanho padrão da chave na Amazon EMR é 256 bits.

Para configurar o HadoopKMS, use a `hadoop-kms-site` classificação para alterar as configurações. Para configurar ACLs, você usa a classificação `kms-acls`.

Para obter mais informações, consulte a documentação do [Hadoop KMS](#). O Hadoop KMS é usado na criptografia transparente do HDFS Hadoop. Para saber mais sobre criptografia HDFS transparente, consulte o tópico sobre [criptografia HDFS transparente](#) na documentação do Apache Hadoop.

Note

Na AmazonEMR, KMS over não HTTPS está habilitado por padrão com o HadoopKMS. Para saber como habilitar o KMS overHTTPS, consulte a documentação do [Hadoop KMS](#).

Important

O Hadoop KMS exige que seus nomes de chave estejam em minúsculas. Se você usar uma chave que tenha caracteres em maiúsculas, seu cluster falhará durante a execução.

Configurando o Hadoop KMS na Amazon EMR

Usando a EMR versão 4.6.0 ou posterior da Amazon, `kms-http-port` é 9700 e `kms-admin-port` 9701.

Você pode configurar o Hadoop KMS no momento da criação do cluster usando a configuração das versões API da AmazonEMR. A seguir estão as classificações de objetos de configuração disponíveis para o Hadoop: KMS

Classificações de configuração do Hadoop KMS

Classificação	Nome do arquivo
hadoop-kms-site	kms-site.xml
hadoop-kms-acls	kms-acls.xml
hadoop-kms-env	kms-env.sh
hadoop-kms-log4j	kms-log4j.properties

Para configurar o Hadoop usando o KMS ACLs CLI

- Crie um cluster com o Hadoop KMS ACLs usando o seguinte comando:

```
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

Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

myConfig.json:

```
[
  {
    "Classification": "hadoop-kms-acls",
    "Properties": {
      "hadoop.kms.blacklist.CREATE": "hdfs,foo,myBannedUser",
      "hadoop.kms.acl.ROLLOVER": "myAllowedUser"
    }
  }
]
```

Para desativar o KMS cache do Hadoop usando o CLI

- Crie um cluster com o Hadoop KMS `hadoop.kms.cache.enable` definido como `false`, usando o seguinte comando:

```
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

Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

myConfig.json:

```
[
  {
    "Classification": "hadoop-kms-site",
    "Properties": {
      "hadoop.kms.cache.enable": "false"
    }
  }
]
```

Para definir variáveis de ambiente no **kms-env.sh** script usando o CLI

- Altere configurações em **kms-env.sh** por meio da configuração **hadoop-kms-env**. Crie um cluster com o Hadoop KMS usando o seguinte comando:

```
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

Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

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": [
    ]
  }
]
}
]

```

[Para obter informações sobre como configurar o HadoopKMS, consulte a documentação do Hadoop. KMS](#)

HDFScriptografia transparente em EMR clusters com vários nós principais

[O Apache Ranger](#) KMS é usado em um EMR cluster da Amazon com vários nós primários para criptografia transparente em. HDFS

O Apache Ranger KMS armazena sua chave raiz e as chaves de zona de criptografia (EZ) em sua Amazon RDS para um EMR cluster da Amazon com vários nós primários. Para habilitar a criptografia transparente HDFS em um EMR cluster da Amazon com vários nós primários, você deve fornecer as seguintes configurações.

- Amazon RDS ou sua própria conexão My SQL server URL para armazenar a chave KMS raiz Ranger e a chave EZ
- Nome de usuário e senha para My SQL
- Senha para a chave KMS raiz do Ranger
- PEMArquivo de Autoridade Certificadora (CA) para SSL conexão com Meu SQL servidor

Você pode fornecer essas configurações usando as classificações `ranger-kms-dbks-site` e `ranger-kms-db-ca`, como demonstra o exemplo a seguir.

```

[
  {
    "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"
  }
}
]

```

A seguir estão as classificações de objetos de configuração para o Apache Ranger. KMS

Classificações de configuração do Hadoop KMS

Classificação	Descrição
ranger-kms-dbks-site	Altere os valores no arquivo dbks-site.xml do Ranger. KMS
ranger-kms-site	Altere os valores no ranger-kms-site arquivo.xml do RangerKMS.
ranger-kms-env	Altere os valores no KMS ambiente Ranger.
ranger-kms-log4j	Altere os valores no arquivo kms-log4j.properties do Ranger. KMS
ranger-kms-db-ca	Altere os valores do arquivo CA no S3 para Minha SQL SSL conexão com o RangerKMS.

Considerações

- É altamente recomendável que você criptografe sua RDS instância da Amazon para melhorar a segurança. Para obter mais informações, consulte [Visão geral da criptografia de RDS recursos da Amazon](#).

- É altamente recomendável que você use Meu SQL banco de dados separado para cada EMR cluster da Amazon com vários nós primários para uma barra de alta segurança.
- Para configurar a criptografia transparente HDFS em um EMR cluster da Amazon com vários nós primários, você deve especificar a `hdfs-encryption-zones` classificação ao criar o cluster. Caso contrário, o Ranger não KMS será configurado ou iniciado. A reconfiguração da `hdfs-encryption-zones` classificação ou de qualquer uma das classificações de KMS configuração do Hadoop em um cluster em execução não é suportada no cluster da Amazon com vários EMR nós primários.

Criar ou executar uma aplicação Hadoop

Tópicos

- [Criar binários usando o Amazon EMR](#)
- [Processamento de dados com streaming](#)
- [Processe dados com um personalizado JAR](#)

Criar binários usando o Amazon EMR

Você pode usar o Amazon EMR como um ambiente de criação para compilar programas que serão usados no cluster. Os programas que você usa com o Amazon EMR devem ser compilados em um sistema que executa a mesma versão do Linux usada pelo Amazon EMR. Para uma versão de 32 bits, você deve compilar em uma máquina de 32 bits ou em uma máquina com as opções de compilação cruzada de 32 bits ativadas. Para uma versão de 64 bits, você deve compilar em uma máquina de 64 bits ou em uma máquina com as opções de compilação cruzada de 64 bits ativadas. Para obter mais informações sobre versões de instâncias do EC2, consulte [Planejar e configurar instâncias do EC2](#) no Guia de gerenciamento do Amazon EMR. As linguagens de programação compatíveis incluem C++, Python e C#.

A tabela a seguir descreve as etapas envolvidas na criação e no teste da sua aplicação usando o Amazon EMR.

Processo para a criação de um módulo

- 1 Conecte-se ao nó principal do seu cluster.
- 2 Copie os arquivos de código-fonte para o nó principal.

- 3 Crie os arquivos binários com as otimizações necessárias.
- 4 Copie os arquivos binários do nó principal para o Amazon S3.

Os detalhes de cada uma dessas etapas são abordados nas seções a seguir.

Para se conectar ao nó principal do cluster

- Siga as instruções em [Conectar-se ao nó principal usando SSH](#) no Guia de gerenciamento do Amazon EMR.

Para copiar os arquivos de código-fonte para o nó principal

1. Coloque os arquivos de origem em um bucket do Amazon S3. Para aprender como criar buckets e como transferir dados para o Amazon S3, consulte o [Guia do usuário do Amazon Simple Storage Service](#).
2. Crie uma pasta no cluster do Hadoop para seus arquivos de código-fonte digitando um comando semelhante ao seguinte:

```
mkdir SourceFiles
```

3. Copie os arquivos de origem do Amazon S3 para o nó principal digitando um comando semelhante ao seguinte:

```
hadoop fs -get s3://mybucket/SourceFiles SourceFiles
```

Crie os arquivos binários com as otimizações necessárias

Como você cria seus arquivos binários depende de vários fatores. Siga as instruções específicas das suas ferramentas de criação para instalar e configurar seu ambiente. Você pode usar os comandos de especificação do sistema do Hadoop para obter as informações do cluster que vão determinar como instalar o seu ambiente de criação.

Para identificar as especificações do sistema

- Use os comandos a seguir para verificar a arquitetura que você está usando para criar seus arquivos binários.

- a. Para visualizar a versão do Debian, insira o seguinte comando:

```
master$ cat /etc/issue
```

A saída será semelhante à seguinte.

```
Debian GNU/Linux 5.0
```

- b. Para visualizar o nome DNS público e o tamanho do processador, insira o seguinte comando:

```
master$ uname -a
```

A saída será semelhante à seguinte.

```
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. Para visualizar a velocidade do processador, insira o seguinte comando:

```
master$ cat /proc/cpuinfo
```

A saída será semelhante à seguinte.

```
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  
...
```

Após a criação dos binários, você pode copiar os arquivos para o Amazon S3.

Copiar os binários do nó principal para o Amazon S3

- Digite o comando a seguir para copiar os binários para o bucket do Amazon S3:

```
hadoop fs -put BinaryFiles s3://mybucket/BinaryDestination
```

Processamento de dados com streaming

O streaming do Hadoop é um utilitário que vem com o Hadoop que permite desenvolver MapReduce executáveis em linguagens diferentes de Java. O streaming é implementado na forma de um JAR arquivo, para que você possa executá-lo na Amazon EMR API ou na linha de comando, como um JAR arquivo padrão.

Esta seção descreve como usar o streaming com a AmazonEMR.

Note

O Apache Hadoop Streaming é uma ferramenta independente. Assim sendo, não descreveremos suas funções e parâmetros aqui. Para obter mais informações sobre o streaming do Hadoop, acesse [http://hadoop.apache.org/docs/stable/hadoop-streaming/HadoopStreaming](http://hadoop.apache.org/docs/stable/hadoop-streaming/HadoopStreaming.html) .html.

Utilização do utilitário Hadoop Streaming

Esta seção descreve como usar o utilitário Hadoop Streaming.

Processo do Hadoop

- 1 Escreva os programas executáveis do mapeador e do reducer na linguagem de programação de sua escolha.

Siga as instruções na documentação do Hadoop para escrever seus executáveis de streaming. Os programas devem ler os dados de entrada de uma entrada padrão e escrever os dados de saída em uma saída padrão. Por padrão, cada linha de entrada/saída representa um registro e o primeiro caractere de tabulação em cada linha é usado como um separador entre a chave e o valor.
- 2 Teste os executáveis localmente e carregue-os para o Amazon S3.
- 3 Use a interface de linha de comando da Amazon ou o EMR console da Amazon para executar seu aplicativo.

Um script de mapeador é executado como um processo separado no cluster. Um executável do reducer transforma a saída de um executável do mapeador em uma saída de dados do fluxo de trabalho.

Os parâmetros `input`, `output`, `mapper` e `reducer` são necessários na maioria das aplicações de streaming. A tabela a seguir descreve esses e outros parâmetros opcionais.

Parâmetro	Descrição	Obrigatório
<code>-input</code>	Local dos dados de entrada no Amazon S3. Tipo: string Padrão: nenhum Restrição:URI. Se não há protocolo especificado, ele usa o sistema de arquivos padrão do cluster.	Sim
<code>-output</code>	Local no Amazon S3 onde a Amazon EMR carrega os dados processados. Tipo: string Padrão: nenhum Restrição: URI Padrão: se um local não for especificado, a Amazon EMR carrega os dados para o local especificado por <code>input</code> .	Sim
<code>-mapper</code>	Nome do executável do mapeador. Tipo: string Padrão: nenhum	Sim
<code>-reducer</code>	Nome do executável do reducer. Tipo: string	Sim

Parâmetro	Descrição	Obrigatório
	Padrão: nenhum	
-cacheFile	<p>Um local no Amazon S3 que contém os arquivos que o Hadoop vai copiar para o diretório de trabalho local (com o objetivo principal de melhorar a performance).</p> <p>Tipo: string</p> <p>Padrão: nenhum</p> <p>Restrições: [URI] # [nome do link simbólico a ser criado no diretório de trabalho]</p>	Não
-cacheArchive	<p>JARarquivo a ser extraído no diretório de trabalho</p> <p>Tipo: string</p> <p>Padrão: nenhum</p> <p>Restrições: [URI] # [nome do diretório do link simbólico a ser criado no diretório de trabalho]</p>	Não
-combiner	<p>Combina os resultados</p> <p>Tipo: string</p> <p>Padrão: nenhum</p> <p>Restrições: nome da classe Java</p>	Não

O código de exemplo a seguir é um executável do mapeador escrito em Python. Esse script faz parte do aplicativo WordCount de amostra.

```
#!/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)
```

Enviar uma etapa de streaming

Esta seção aborda os conceitos básicos do envio de uma etapa de streaming para um cluster. Uma aplicação de streaming lê os dados de entrada de uma entrada padrão e, em seguida, executa um script ou executável (denominado mapeador) para cada entrada. O resultado de cada uma das entradas é salvo localmente, normalmente em uma partição Hadoop Distributed File System (HDFS). Depois que todas as entradas são processadas pelo mapeador, um segundo script ou programa executável (chamado de reducer) processa os resultados do mapeador. Os resultados do reducer são enviados para a saída padrão. Você pode encadear uma série de etapas de streaming, em que a saída de uma etapa se torna a entrada de outra etapa.

O mapeador e o reducer podem ser referenciados como arquivos ou você pode fornecer uma classe Java. Você pode implementar o mapeador e o redutor em qualquer uma das linguagens suportadas, incluindo Ruby, Perl, PHP Python ou Bash.

Enviar uma etapa de streaming usando o console

Este exemplo descreve como usar o EMR console da Amazon para enviar uma etapa de streaming para um cluster em execução.

Enviar uma etapa de streaming

1. Abra o EMR console da Amazon em <https://console.aws.amazon.com/emr>.
2. Em Cluster List (Lista de clusters), selecione o nome do cluster.
3. Role até a seção Steps (Etapas) e expanda-a. Em seguida, escolha Add step (Adicionar etapa).
4. Na caixa de diálogo Add Step (Adicionar etapa):
 - Em Step type (Tipo de etapa), escolha Streaming program (Programa de streaming).

- Em Name (Nome), aceite o nome padrão (programa Streaming) ou digite um novo nome.
 - Em Mapper, digite ou navegue até o local da classe do mapeador no Hadoop ou de um bucket do S3, onde se encontra o executável do mapeador como, por exemplo um programa Python. O valor do caminho deve estar no formato *BucketName/path/MapperExecutable*.
 - Em Reducer (Redutor), digite ou navegue até o local da classe do reducer no Hadoop ou de um bucket do S3 onde se encontra o executável do reducer, por exemplo um programa Python. O valor do caminho deve estar no formato *BucketName/path/ReducerExecutable*. A Amazon EMR oferece suporte à palavra-chave agregada especial. Para obter mais informações, acesse a biblioteca Aggregate fornecida pelo Hadoop.
 - Em Input S3 location (Localização do S3 de entrada), digite ou navegue até o local dos dados de entrada.
 - Em Local de saída do S3, digite ou navegue até o nome do bucket de saída do Amazon S3.
 - Para Arguments (Argumentos), deixe o campo em branco.
 - Para Action on failure (Ação na falha), aceite a opção padrão Continue (Continuar).
5. Escolha Adicionar. A etapa é exibida no console com o status Pendente.
 6. O status da etapa muda de Pending (Pendente) para Running (Em execução) e depois para Completed (Concluído) conforme ela é executada. Para atualizar o status, escolha o ícone Refresh (Atualizar) acima da coluna Actions (Ações).

AWS CLI

Esses exemplos demonstram como usar o AWS CLI para criar um cluster e enviar uma etapa de streaming.

Para criar um cluster e enviar uma etapa de streaming usando o AWS CLI

- Para criar um cluster e enviar uma etapa de streaming usando o AWS CLI, digite o comando a seguir e substitua *myKey* com o nome do seu EC2 key pair. Observe que o argumento para `--files` deve ser o caminho do Amazon S3 para o local do script e os argumentos para `-mapper` e `-reducer` devem ser os nomes dos respectivos arquivos de script.

```
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

Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

Quando você especifica a contagem de instâncias sem usar o parâmetro `--instance-groups`, um único nó principal é executado, e as instâncias restantes são executadas como nós core. Todos os nós usam o tipo de instância especificado no comando.

Note

Se você ainda não criou a função de EMR serviço e o perfil de EC2 instância padrão da Amazon, digite `aws emr create-default-roles` para criá-los antes de digitar o `create-cluster` subcomando.

Para obter mais informações sobre o uso dos EMR comandos da Amazon no AWS CLI, consulte <https://docs.aws.amazon.com/cli/latest/reference/emr>.

Processe dados com um personalizado JAR

Um customizado JAR executa um programa Java compilado que você pode carregar no Amazon S3. Você deve compilar o programa em relação à versão do Hadoop que deseja iniciar e enviar uma CUSTOM_JAR etapa para o seu cluster da Amazon. EMR Para obter mais informações sobre como compilar um JAR arquivo, consulte [Criar binários usando o Amazon EMR](#).

Para obter mais informações sobre a criação de um MapReduce aplicativo Hadoop, consulte o [MapReduce Tutorial](#) na documentação do Apache Hadoop.

Tópicos

- [Envie uma JAR etapa personalizada](#)

Envie uma JAR etapa personalizada

Um customizado JAR executa um programa Java compilado que você pode carregar no Amazon S3. Você deve compilar o programa em relação à versão do Hadoop que deseja iniciar e enviar uma CUSTOM_JAR etapa para o seu cluster da Amazon. EMR Para obter mais informações sobre como compilar um JAR arquivo, consulte [Criar binários usando o Amazon EMR](#).

Para obter mais informações sobre a criação de um MapReduce aplicativo Hadoop, consulte o [MapReduce Tutorial](#) na documentação do Apache Hadoop.

Esta seção aborda os princípios básicos do envio de uma JAR etapa personalizada na Amazon. EMR O envio de uma JAR etapa personalizada permite que você escreva um script para processar seus dados com a linguagem de programação Java.

Envie uma JAR etapa personalizada com o console

Este exemplo descreve como usar o EMR console da Amazon para enviar uma JAR etapa personalizada para um cluster em execução.

Para enviar uma JAR etapa personalizada com o console

1. Abra o EMR console da Amazon em <https://console.aws.amazon.com/emr>.
2. Em Cluster List (Lista de clusters), selecione o nome do cluster.
3. Role até a seção Steps (Etapas) e expanda-a. Em seguida, escolha Add step (Adicionar etapa).
4. Na caixa de diálogo Add Step (Adicionar etapa):
 - Em Tipo de etapa, escolha Personalizado JAR.
 - Em Nome, aceite o nome padrão (PersonalizadoJAR) ou digite um novo nome.
 - Para a localização do JAR S3, digite ou navegue até a localização do seu JAR arquivo. JARlocalização talvez seja um caminho para o S3 ou uma classe java totalmente qualificada no caminho de classe..
 - Para Arguments (Argumentos), digite argumentos necessários como strings separadas por espaços ou deixe o campo em branco.
 - Para Action on failure (Ação na falha), aceite a opção padrão Continue (Continuar).
5. Escolha Adicionar. A etapa é exibida no console com o status Pendente.
6. O status da etapa muda de Pending (Pendente) para Running (Em execução) e depois para Completed (Concluído) conforme ela é executada. Para atualizar o status, escolha o ícone Refresh (Atualizar) acima da coluna Actions (Ações).

Iniciando um cluster e enviando uma JAR etapa personalizada com o AWS CLI

Para iniciar um cluster e enviar uma JAR etapa personalizada com o AWS CLI

Para iniciar um cluster e enviar uma JAR etapa personalizada com o AWS CLI, digite o `create-cluster` subcomando com o `--steps` parâmetro.

- Para iniciar um cluster e enviar uma JAR etapa personalizada, digite o seguinte comando, substitua *myKey* com o nome do seu EC2 key pair e substitua *mybucket* com o nome do seu bucket.

```
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

Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

Quando você especifica a contagem de instâncias sem usar o parâmetro `--instance-groups`, um único nó primário é iniciado e as instâncias restantes são iniciadas como nós centrais. Todos os nós usam o tipo de instância que você especifica no comando.

Note

Se você ainda não criou a função de EMR serviço e o perfil de EC2 instância padrão da Amazon, digite `aws emr create-default-roles` para criá-los antes de digitar o `create-cluster` subcomando.

Para obter mais informações sobre o uso dos EMR comandos da Amazon no AWS CLI, consulte <https://docs.aws.amazon.com/cli/latest/reference/emr>.

Dependências de terceiros

Às vezes, pode ser necessário incluir no MapReduce classpath JARs para uso com seu programa. Existem duas opções para fazer isso:

- Inclua o `--libjars s3://URI_to_JAR` nas opções da etapa para o procedimento em [Iniciando um cluster e enviando uma JAR etapa personalizada com o AWS CLI](#).
- Inicie o cluster com uma configuração `mapreduce.application.classpath` modificada em `mapred-site.xml`. Use a classificação para a configuração `mapred-site`. Para criar o cluster com a etapa usando AWS CLI, isso teria a seguinte aparência:

```
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"
    }
  }
]
```

A lista de caminhos separados por vírgula deve ser anexada ao caminho de JVM classe de cada tarefa.

Leia objetos restaurados

Com a EMR versão 7.2.0 e superior da Amazon, você pode ler objetos Glacier restaurados na localização S3 da tabela com o protocolo. S3A Os mecanismos das versões anteriores não fazem distinção entre arquivos Glacier e Glacier Deep Archive, o que significa que você obterá um `AmazonS3Exception` se tentasse acessar um arquivo Glacier em

andamento com. S3A Essa operação de leitura ignora os arquivos arquivados do Glacier se eles ainda estiverem sendo restaurados. Para ativar esse comportamento, use a configuração `fs.s3a.glacier.read.restored.objects`. Essa configuração pode ter três valores:

- `READ_ALL` — esse valor indica que a Amazon não EMR deve contabilizar as classes de armazenamento recuperadas do Amazon S3. Esse é o comportamento padrão.
- `SKIP_ALL_GLACIER` — esse valor indica que a Amazon EMR deve ignorar todos os objetos do S3 que estejam marcados com a classe de armazenamento Glacier e recuperar todos os outros objetos. Esse é o comportamento padrão do Amazon Athena em relação aos objetos Glacier.
- `READ_RESTORED_GLACIER_OBJECTS` — esse valor indica que a Amazon EMR deve verificar o status restaurado do objeto Glacier. Se a Amazon EMR puder restaurar o objeto, você poderá lê-lo como um objeto S3 normal. Caso contrário, a Amazon EMR ignorará o objeto do S3 Glacier

Exemplos

Spark

Para ler objetos restaurados ao usar o Spark, use a seguinte configuração:

```
--conf spark.hadoop.fs.s3a.glacier.read.restored.objects=<value>
```

Se você usa spark-sql, use a seguinte configuração em vez disso;

```
spark-sql --conf spark.hadoop.fs.s3a.glacier.read.restored.objects=<value>
```

Flink

Se você usa o Flink, pode definir a configuração no arquivo `flink-conf.yaml`

```
fs.s3a.glacier.read.restored.objects: <value>
```

Você também pode definir a `flink-conf` classificação:

```
[  
  {
```

```
"Classification": "flink-conf",
"Properties": {
  "fs.s3a.glacier.read.restored.objects": "<value>"
}
]
```

Hive

Se você usa o Hive, defina a configuração no `hive-site.xml` arquivo.

```
<property>
  <name>fs.s3a.glacier.read.restored.objects</name>
  <value><value></value>
</property>
```

Você também pode usar o Hive CLI para definir a propriedade `--hiveconf`:

```
hive --hiveconf fs.s3a.glacier.read.restored.objects=<value>
```

Considerações

Ao ler objetos restaurados do S3 Glacier, observe as seguintes considerações:

- Você só pode ler objetos restaurados se usar o S3A esquema ou o S3AFileSystem para acessar os dados.
- Quando você lê um objeto Glacier restaurado, a Amazon EMR não restaura o objeto em si. Para fazer isso, você deve usar o AWS CLI ou AWS SDK o.

Ative o reconhecimento de acesso não uniforme à memória para YARN contêineres

Com EMR as versões 6.x e posteriores da Amazon, você pode usar o acesso não uniforme à memória (NUMA) para multiprocessar seus dados em clusters. NUMA é um padrão de design de memória de computador em que o processador pode acessar sua própria memória local mais rapidamente do que a memória em outro processador ou compartilhada entre processadores. YARN os contêineres têm melhor desempenho NUMA porque podem se vincular a um NUMA nó

específico que atende a todas as alocações de memória subsequentes. Isso reduz a quantidade de vezes que o cluster precisa acessar a memória remota.

Você pode ativar o NUMA suporte para YARN contêiner quando a máquina do nó de trabalho for um NUMA multinó. Para confirmar se o nó de processamento é um nó NUMA único ou multi-NUMA, execute o comando a seguir.

```
lscpu | grep -i numa
NUMA node(s): 2
```

Em geral, instâncias superiores a 12x têm dois nós NUMA. Isso não se aplica a instâncias metal.

Para ativar a NUMA conscientização sobre YARN contêineres

1. Use a seguinte `yarn-site` configuração em seu 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. Forneça a ação de bootstrap a seguir no 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. Cada contêiner deve ter reconhecimento de NUMA. Você pode notificar a máquina virtual Java (JVM) em cada contêiner com um NUMA sinalizador. Por exemplo, para notificar o JVM para uso NUMA em um MapReduce trabalho, adicione as seguintes propriedades em `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. Para verificar se você ativou o NUMA, pesquise algum dos arquivos de log do NodeManager com o comando a seguir.

```
grep "NUMA resources allocation is enabled," *
```

Para verificar se NodeManager atribuiu recursos de NUMA nó a um contêiner, pesquise o NodeManager log com o comando a seguir, substituindo-o `<container_id>` pelo seu próprio ID de contêiner.

```
grep "NUMA node" | grep <container_id>
```


Histórico de versões do Hadoop

A tabela a seguir lista a versão do Hadoop incluída em cada versão da AmazonEMR, junto com os componentes instalados com o aplicativo. Para obter as versões dos componentes em cada versão, consulte a seção Versão de componentes da sua versão em [Versões de lançamento do Amazon EMR 7.x](#), [Versões de lançamento do Amazon EMR 6.x](#) ou [Versões de lançamento do Amazon EMR 5.x](#).

Informações de versão do Hadoop

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o 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-nodemanager, 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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
emr-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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
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
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

Gravadora Amazon EMR Release	Versão do Hadoop	Componentes instalados com o Hadoop
emr-4.0.0	2.6.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-mapred, hadoop-hdfs-datano de, hadoop-hdfs-namenode, hadoop-httpfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager

Notas de versão do Hadoop por versão

Tópicos

- [Amazon EMR 7.2.0 - Notas de lançamento do Hadoop](#)
- [Amazon EMR 6.6.0 - Notas de lançamento do Hadoop](#)

Amazon EMR 7.2.0 - Notas de lançamento do Hadoop

Amazon EMR 7.2.0 - Mudanças no Hadoop

Tipo	Descrição
Novo recurso	HADOOP-18850 : S3A: habilite a criptografia de camada dupla do lado do servidor com chaves AWS KMS
Novo recurso	Support S3 Glacier leia objetos restaurados
Melhoria	Escalando dinamicamente as conexões máximas do S3A com base nos núcleos ao executar Spark Jobs

Tipo	Descrição
Melhoria	HADOOP-18797 : Support Concurrent Writes com o S3A Magic Committer
Melhoria	HADOOP-19047 : Support InMemory Tracking of S3A Magic Commits
Melhoria	Adicione o método "Exportar", "Publicar" e "AbortPublish" no MagicV2 Committer
Melhoria	Implementar ReadFullyIntoBuffers FileSystem API no S3A
Correção de bugs	HADOOP-18793 : O S3A StagingCommitter não limpa o diretório de staging-uploads
Upgrade	HADOOP-19008 : S3A: atualize a versão aws-sdk para 2.21.41
Upgrade	HADOOP-18613 : Atualize ZooKeeper para a versão 3.8.3

Amazon EMR 7.2.0 - Recursos do Hadoop

Veja a lista a seguir para ver os novos recursos do Hadoop na Amazon EMR 7.2.0.

- [Suporte ao S3 Glacier para ler objetos restaurados](#) — Com a EMR versão 7.2.0 e superior da Amazon, você pode ler objetos Glacier restaurados a partir da localização S3 da tabela com o protocolo. S3A Essa operação de leitura ignora os arquivos arquivados do Glacier se eles ainda estiverem sendo restaurados.
- Support gravações simultâneas com o commit mágico do S3A — O EMR Amazon 7.2.0 introduz suporte para gravações simultâneas com o commit mágico do S3A. [Esse recurso foi adicionado à comunidade OSS Hadoop como parte do HADOOP -18797.](#)

Amazon EMR 6.6.0 - Notas de lançamento do Hadoop

Amazon EMR 6.6.0 - Mudanças no Hadoop

Tipo	Descrição
Bug	Registros duplicados corrigidos ao ler arquivos BZip2 de texto.
Backport	HADOOP-18136 : Verifique o FileUtils .unTar() tratamento de arquivos.tar ausentes
Backport	HADOOP-17627 : Backport para a ramificação 3.2 HADOOP -17371, -17621, -17625 para atualizar o Jetty para 9.4.39 HADOOP HADOOP
Backport	HADOOP-17655 : Atualize o Jetty para 9.4.40
Backport	HADOOP-17796 : Atualize a versão jetty para 9.4.43
Backport	HADOOP-17661 : versões mvn:set falha ao analisar pom.xml
Backport	HADOOP-17236 : aumente snakeyaml para 1,26 para mitigar -2017-18640 CVE
Backport	HADOOP-16717 : Remova a dependência 4 do GenericsUtil isLog Log4 jLogger jLoggerAdapter
Backport	HADOOP-17633 : aumente o json-smart para 2.4.2 e 9.8 devido a nimbus-jose-jwt CVEs
Backport	HADOOP-17844 : Atualize JSON o smart para 2.4.7

Tipo	Descrição
Backport	HADOOP-17972 : Backport HADOOP -17683 (atualização commons-io para 2.8.0) para branch-3.2
Backport	HADOOP-16555 : Atualize o commons-compress para 1.19
Backport	HADOOP-17370 : Atualize o commons-compress para 1.21
Backport	HADOOP-17096 : Corrige ZStandard Compressor o deslocamento do buffer de entrada
Backport	HADOOP-17112 : Não é permitido espaço em branco nos caminhos ao salvar arquivos no s3a via committer
Backport	HADOOP-13500 : Sincronizando a iteração do objeto de propriedades de configuração
Backport	HDFS-14099 : Descritor de quadro desconhecido ao descompactar vários quadros em ZStandardDecompressor
Backport	HDFS-16410 : Análise insegura de Xml em OfflineEditsXMLLoader
Backport	HDFS-14498 : LeaseManager pode repetir para sempre o arquivo cuja criação falhou
Backport	HDFS-15290 : NPE entrada durante a inicialização HttpServer NameNode
Backport	HDFS-15293 : Relaxe a condição para aceitar uma fsimage ao receber um posto de controle

Tipo	Descrição
Backport	HDFS-12979 : StandbyNode deve fazer o upload FsImage para depois do checkpoint ObserverNode
Backport	YARN-10538 : Adicionar nós de recomissio namento à lista de nós atualizados retornados ao AM
Backport	YARN-10472 : Backport YARN -10314 (lançado apenas com jars de cliente YarnClient sombreados) NoClassDefFoundError para WebSocketException a ramificação-3.2
Backport	YARN-9968 : O Public Localizer está saindo devido a NodeManager NullPointerException
Backport	YARN-10651 : CapacityScheduler caiu com o in NPE AbstractYarnScheduler.updateNodeResource()
Backport	YARN-9339 : Métrica de aplicativos pendentes incorreta após mover o aplicativo para uma nova fila
Backport	YARN-10438 : manipular nulo em containerId ClientRMService#getContainerReport()
Backport	YARN-7266 : ATS 1.5 falha ao iniciar se os RollingLevelDb arquivos estiverem corrompidos ou ausentes
Backport	YARN-9063 : ATS 1.5 falha ao iniciar se os RollingLevelDb arquivos estiverem corrompidos ou ausentes

Tipo	Descrição
Backport	YARN-9848 : Reverta YARN -4946 (o RM não deve considerar um aplicativo como COMPLETED quando a agregação de registros não está em um estado terminal).

Apache HBase

[HBase](#) é um banco de dados distribuído, não relacional e de código aberto desenvolvido como parte do projeto Hadoop da Apache Software Foundation. HBase é executado sobre o Hadoop Distributed File System (HDFS) para fornecer recursos de banco de dados não relacional para o ecossistema Hadoop. HBase está incluído na EMR versão 4.6.0 e posterior da Amazon.

HBase funciona perfeitamente com o Hadoop, compartilhando seu sistema de arquivos e servindo como entrada e saída diretas para a MapReduce estrutura e o mecanismo de execução. HBase também se integra ao Apache Hive, permitindo consultas SQL semelhantes a HBase tabelas, uniões com tabelas baseadas em Hive e suporte para conectividade de banco de dados Java (JDBC). Para obter mais informações sobre HBase, consulte o [Apache HBase](#) e a [HBase documentação](#) no site do Apache. Para ver um exemplo de como usar HBase com o Hive, consulte a postagem do blog de AWS Big Data [Combine No SQL e análises massivamente paralelas usando o Apache HBase e o Apache Hive na Amazon](#). EMR

Com o HBase on Amazon EMR, você também pode fazer backup de seus HBase dados diretamente no Amazon Simple Storage Service (Amazon S3) e restaurar a partir de um backup criado anteriormente ao iniciar um cluster. HBase A Amazon EMR oferece opções adicionais de integração com o Amazon S3 para persistência de dados e recuperação de desastres.

- HBase no Amazon S3 - Com a EMR versão 5.2.0 e posterior da Amazon, você pode usar no Amazon HBase S3 para armazenar o diretório HBase raiz e os metadados de um cluster diretamente no Amazon S3. Posteriormente, você pode iniciar um novo cluster, apontando-o para o local do diretório de raiz no Amazon S3. Somente um cluster por vez pode usar a HBase localização no Amazon S3, com exceção de um cluster de réplica de leitura. Para obter mais informações, consulte [HBase no Amazon S3 \(modo de armazenamento Amazon S3\)](#).
- HBase réplicas de leitura - A Amazon EMR versão 5.7.0 e posterior com o Amazon HBase S3 oferece suporte a clusters de réplicas de leitura. Um cluster de réplica de leitura fornece acesso somente leitura aos arquivos de armazenamento de um cluster primário e a metadados para operações somente leitura. Para obter mais informações, consulte [Usar um cluster de réplica de leitura](#).
- HBase Snapshots - Como alternativa ao HBase Amazon S3, EMR com a versão 4.0 e posterior, você pode criar instantâneos de HBase seus dados diretamente no Amazon S3 e depois recuperar dados usando os snapshots. Para obter mais informações, consulte [Usando HBase instantâneos](#).

⚠ Important

Para a escalabilidade de EMR HBase clusters da Amazon, não recomendamos usar [escalabilidade gerenciada ou escalabilidade com políticas personalizadas com clusters](#).

HBase

A tabela a seguir lista a versão HBase incluída na versão mais recente da série Amazon EMR 7.x, junto com os componentes com os quais a Amazon é EMR instalada. HBase

Para a versão dos componentes instalados HBase nesta versão, consulte [Versões de componentes da versão 7.2.0](#).

HBase informações sobre a versão do emr-7.2.0

Gravadora EMR de lançamento da Amazon	HBase Versão	Componentes instalados com o 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-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

A tabela a seguir lista a versão HBase incluída na versão mais recente da série EMR 6.x da Amazon, junto com os componentes com os quais a Amazon é EMR instalada. HBase

Para a versão dos componentes instalados HBase nesta versão, consulte [Versões de componentes da versão 6.15.0](#).

HBase informações da versão do emr-6.15.0

Gravadora EMR de lançamento da Amazon	HBase Versão	Componentes instalados com o 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-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

Note

O Apache HBase HBCK2 é uma ferramenta operacional separada para reparar HBase regiões e tabelas do sistema. Na Amazon EMR versão 6.1.0 e posterior, o `hbase-hbck2.jar` é fornecido `/usr/lib/hbase-operator-tools/` no nó primário. Para obter mais informações sobre como criar e usar a ferramenta, consulte [HBase HBCK2](#).

A tabela a seguir lista a versão HBase incluída na versão mais recente da série Amazon EMR 5.x, junto com os componentes com os quais a Amazon é EMR instalada. HBase

Para a versão dos componentes instalados HBase nesta versão, consulte [Versões de componentes da versão 5.36.2](#).

HBase informações da versão do emr-5.36.2

Gravadora EMR de lançamento da Amazon	HBase Versão	Componentes instalados com o 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Tópicos

- [Criar um cluster com o HBase](#)
- [HBase no Amazon S3 \(modo de armazenamento Amazon S3\)](#)
- [Registros de gravação antecipada \(WAL\) para Amazon EMR](#)
- [Usando a HBase concha](#)
- [Acesse HBase tabelas com o Hive](#)
- [Usando HBase instantâneos](#)
- [Configurar HBase](#)
- [Visualize a interface HBase do usuário](#)
- [Exibir arquivos de log do HBase](#)
- [Monitor HBase com Ganglia](#)

- [Migrando de versões anteriores HBase](#)
- [HBase histórico de lançamentos](#)

Criar um cluster com o HBase

Os procedimentos desta seção abrangem as noções básicas sobre o lançamento de um cluster usando o AWS Management Console e o AWS CLI. Para obter informações detalhadas sobre como planejar, configurar e lançar EMR clusters da Amazon, consulte [Planejar e configurar clusters](#) no Guia EMR de gerenciamento da Amazon.

Criação de um cluster HBase usando o console

Para obter etapas rápidas para iniciar clusters com o console, consulte [Introdução à Amazon EMR](#) no Guia EMR de gerenciamento da Amazon.

Para iniciar um cluster com HBase instalado usando o console

1. Abra o EMR console da Amazon em <https://console.aws.amazon.com/emr>.
2. Escolha Create cluster (Criar cluster) e Go to advanced options (Ir para opções avançadas).
3. Para Software Configuration (Configuração do software), escolha um valor de 4.6.0 ou posterior para Amazon Release Version (Versão da Amazon) (recomendamos a versão mais recente). Escolha HBase outros aplicativos conforme desejado.
4. Com a Amazon EMR versão 5.2.0 e posterior, em Configurações HBase de armazenamento, selecione HDFS ou S3. Para obter mais informações, consulte [HBase no Amazon S3 \(modo de armazenamento Amazon S3\)](#).
5. Selecione outras opções conforme necessário e escolha Create cluster (Criar cluster).

Criação de um cluster com HBase o uso do AWS CLI

Use o comando a seguir para criar um cluster com HBase instalado:

```
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

Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

Se você usa HBase no Amazon S3, especifique a `--configurations` opção com uma referência a um objeto de JSON configuração. O objeto de configuração deve conter uma `hbase-site` classificação que especifique o local no Amazon S3 HBase onde os dados são armazenados usando `hbase.rootdir` a propriedade. Ele também deve conter uma classificação `hbase`, que especifica `s3` usando a propriedade `hbase.emr.storageMode`. O exemplo a seguir demonstra um JSON trecho com essas configurações.

```
[
  {
    "Classification": "hbase-site",
    "Properties": {
      "hbase.rootdir": "s3://MyBucket/MyHBaseStore"
    }
  },
  {
    "Classification": "hbase",
    "Properties": {
      "hbase.emr.storageMode": "s3"
    }
  }
]
```

Para obter mais informações HBase sobre o Amazon S3, consulte [HBaseno Amazon S3 \(modo de armazenamento Amazon S3\)](#) Para obter mais informações sobre classificações, consulte [Configurar aplicações](#).

HBaseno Amazon S3 (modo de armazenamento Amazon S3)

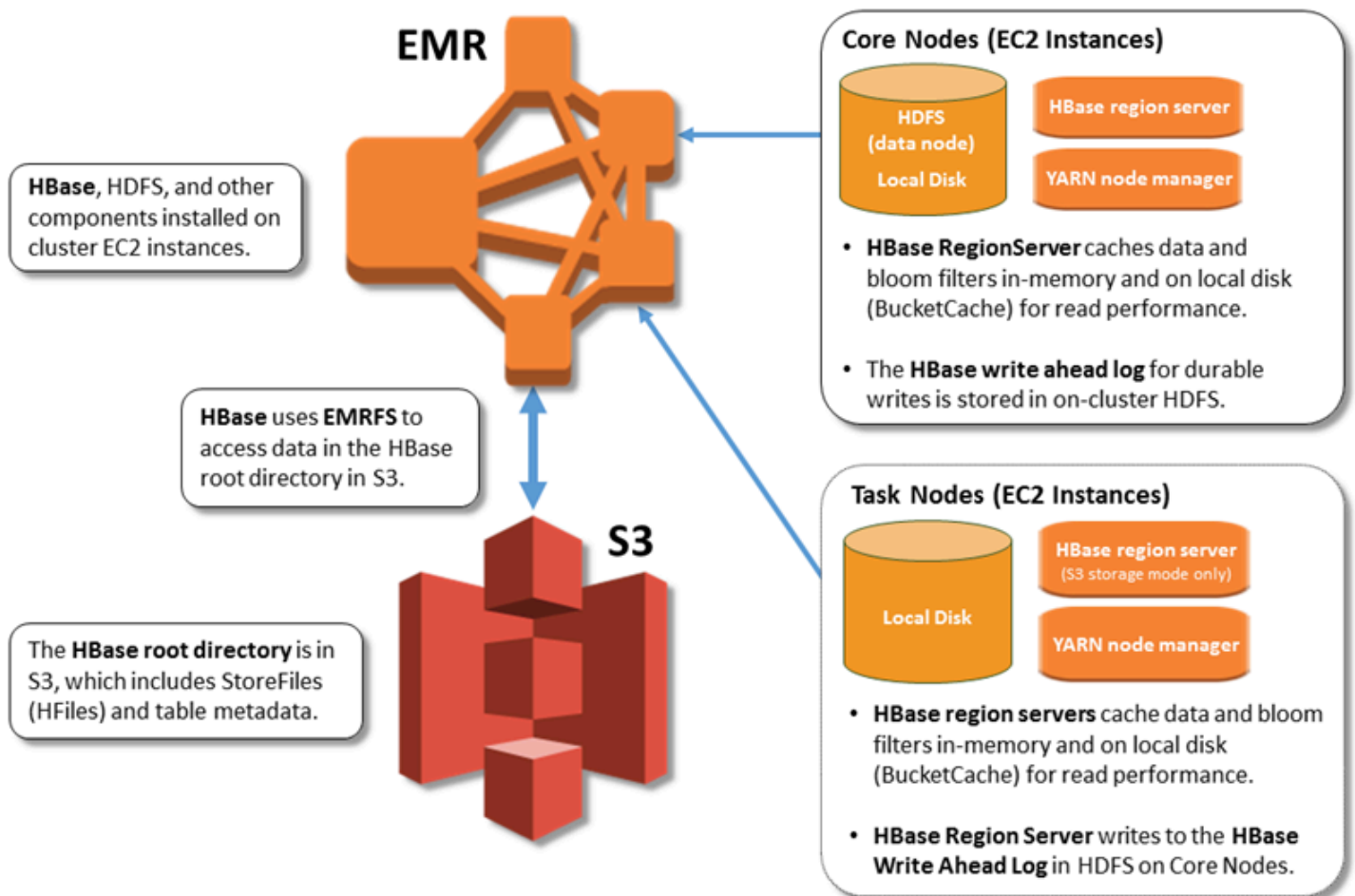
Ao executar HBase na Amazon EMR versão 5.2.0 ou posterior, você pode habilitá-la HBase no Amazon S3, que oferece as seguintes vantagens:

- O diretório HBase raiz é armazenado no Amazon S3, incluindo arquivos de HBase armazenamento e metadados de tabelas. Esses dados são persistentes fora do cluster,

disponíveis nas zonas de EC2 disponibilidade da Amazon, e você não precisa se recuperar usando snapshots ou outros métodos.

- Com arquivos armazenados no Amazon S3, você pode dimensionar seu EMR cluster da Amazon de acordo com seus requisitos de computação em vez de requisitos de dados, com replicação de 3 vezes. HDFS
- Usando a Amazon EMR versão 5.7.0 ou posterior, você pode configurar um cluster de réplica de leitura, que permite manter cópias somente para leitura dos dados no Amazon S3. Você pode acessar os dados do cluster de réplica de leitura para realizar operações de leitura simultaneamente e no evento de o cluster primário se tornar indisponível.
- Na Amazon EMR versão 6.2.0 e posterior, o HFile rastreamento persistente usa uma tabela HBase do sistema chamada `hbase:storefile` para rastrear diretamente os HFile caminhos usados para operações de leitura. Esse atributo é habilitado por padrão e não exige que a migração manual seja executada.

A ilustração a seguir mostra os HBase componentes relevantes HBase no Amazon S3.



Habilitando HBase no Amazon S3

Você pode habilitar HBase no Amazon S3 usando o EMR console da Amazon AWS CLI, o ou o Amazon. EMR API A configuração é uma opção durante a criação do cluster. Ao usar o console, você escolhe a configuração usando Advanced options (Opções avançadas). Ao usar o AWS CLI, use a `--configurations` opção para fornecer um objeto JSON de configuração. As propriedades do objeto de configuração especificam o modo de armazenamento e o local do diretório raiz no Amazon S3. A localização do Amazon S3 que você especificar deve estar na mesma região do seu cluster da AmazonEMR. Somente um cluster ativo por vez pode usar o mesmo diretório HBase raiz no Amazon S3. Para ver as etapas do console e um exemplo detalhado de criação de cluster usando o AWS CLI, consulte. [Criar um cluster com o HBase](#) Um exemplo de objeto de configuração é mostrado no JSON trecho a seguir.

```
{
  "Classification": "hbase-site",
  "Properties": {
    "hbase.rootdir": "s3://my-bucket/my-hbase-rootdir"
  },
}
{
  "Classification": "hbase",
  "Properties": {
    "hbase.emr.storageMode": "s3"
  }
}
```

Note

Se você usar um bucket do Amazon S3 como formaHBase, deverá adicionar uma barra no final do Amazon S3. `rootdir` URI Por exemplo, é necessário usar `"hbase.rootdir: s3://my-bucket/"`, em vez de `"hbase.rootdir: s3://my-bucket"`, para evitar problemas.

Usar um cluster de réplica de leitura

Depois de configurar um cluster primário usando o HBase Amazon S3, você pode criar e configurar um cluster de réplica de leitura que fornece acesso somente de leitura aos mesmos dados do

cluster primário. Isso é útil quando você precisa de acesso simultâneo para consultar dados ou de acesso ininterrupto caso o cluster primário se torne indisponível. O recurso de leitura e réplica está disponível com a Amazon EMR versão 5.7.0 e posterior.

O cluster primário e o cluster de réplica de leitura são configuradas da mesma maneira, com uma diferença importante. Ambos apontam para o mesmo local `hbase.rootdir`. No entanto, a classificação `hbase` para o cluster de réplica de leitura inclui a propriedade `"hbase.emr.readreplica.enabled": "true"`.

Por exemplo, dada a JSON classificação do cluster primário, conforme mostrado anteriormente no tópico, a configuração de um cluster de réplica de leitura é a seguinte:

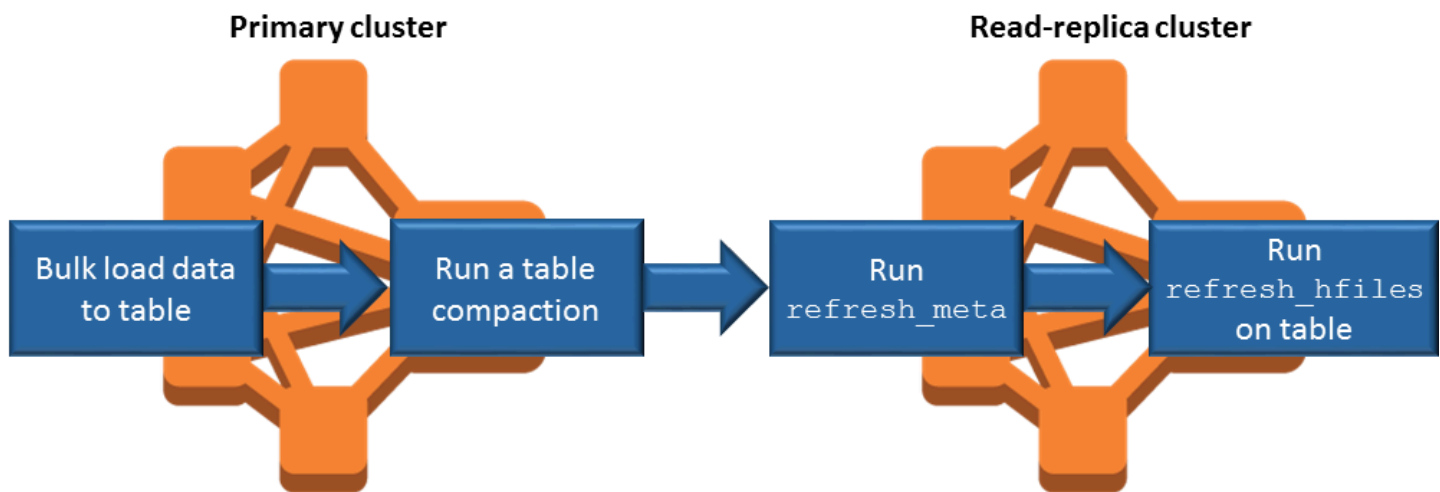
```
{
  "Classification": "hbase-site",
  "Properties": {
    "hbase.rootdir": "s3://my-bucket/my-hbase-rootdir"
  },
  {
    "Classification": "hbase",
    "Properties": {
      "hbase.emr.storageMode": "s3",
      "hbase.emr.readreplica.enabled": "true"
    }
  }
}
```

Sincronizar a réplica de leitura ao adicionar dados

Como a réplica de leitura usa HBase StoreFiles e os metadados que o cluster primário grava no Amazon S3, a réplica de leitura é tão atual quanto o armazenamento de dados do Amazon S3. A orientação a seguir pode ajudar a minimizar o tempo de retardo entre o cluster primário e a réplica de leitura quando você grava dados.

- Carregue os dados em massa no cluster primário sempre que possível. Para obter mais informações, consulte [Carregamento em massa](#) na HBase documentação do Apache.
- Uma liberação que grava arquivos de armazenamento no Amazon S3 deve ocorrer logo que possível após a adição dos dados. Faça a liberação manualmente ou ajuste configurações de liberação para minimizar o tempo de retardo.

- Se compactações puderem ser executadas automaticamente, execute uma compactação manual para evitar inconsistências quando compactações forem acionadas.
- No cluster de réplica de leitura, quando algum metadado for alterado, por exemplo, quando ocorrerem divisões ou compactações de HBase regiões, ou quando tabelas forem adicionadas ou removidas, execute o comando `refresh_meta`
- No cluster de réplica de leitura, execute o comando `refresh_hfiles` quando registros forem adicionados ou alterados em uma tabela.



HFileMonitoramento persistente

O HFile rastreamento persistente usa uma tabela HBase do sistema chamada `hbase:storefile` para rastrear diretamente os HFile caminhos usados para operações de leitura. Novos HFile caminhos são adicionados à tabela à medida que dados adicionais são adicionados HBase. Isso remove as operações de renomeação como um mecanismo de confirmação nas HBase operações críticas do caminho de gravação e melhora o tempo de recuperação ao abrir uma HBase região lendo a tabela do `hbase:storefile` sistema em vez da listagem de diretórios do sistema de arquivos. Esse recurso é ativado por padrão na Amazon EMR versão 6.2.0 e posterior e não exige nenhuma etapa de migração manual.

Note

O HFile rastreamento persistente usando a tabela do sistema HBase `storefile` não oferece suporte ao recurso de replicação da HBase região. Para obter mais informações sobre a

replicação HBase da região, consulte Leituras de alta disponibilidade [consistentes com a linha do tempo](#).

Desativando o rastreamento persistente HFile

O HFile rastreamento persistente é ativado por padrão a partir da EMR versão 6.2.0 da Amazon. Para desativar o HFile rastreamento persistente, especifique a seguinte substituição de configuração ao iniciar um cluster:

```
{
  "Classification": "hbase-site",
  "Properties": {
    "hbase.storefile.tracking.persist.enabled": "false",

    "hbase.hstore.engine.class": "org.apache.hadoop.hbase.regionserver.DefaultStoreEngine"
  }
}
```

Note

Ao reconfigurar o EMR cluster da Amazon, todos os grupos de instâncias devem ser atualizados.

Sincronizar a tabela Storefile manualmente

A tabela de arquivos de armazenamento é mantida atualizada à medida que novos HFiles são criados. No entanto, se a tabela storefile ficar fora de sincronia com os arquivos de dados por qualquer motivo, estes comandos poderão ser usados para sincronizar os dados manualmente:

Sincronizar a tabela storefile em uma região on-line:

```
hbase org.apache.hadoop.hbase.client.example.RefreshHFilesClient <table>
```

Sincronizar a tabela do storefile em uma região off-line:

- Remova o znode da tabela storefile.

```
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]
```

- Atribua a região (execute em “hbase shell”).

```
hbase cli> assign '<region name>'
```

- Se a tarefa falhar.

```
hbase cli> disable '<table name>'
hbase cli> enable '<table name>'
```

Escalar a tabela storefile

Por padrão, a tabela storefile é dividida em quatro regiões. Se a tabela storefile continuar sob carga de gravação pesada, a tabela ainda poderá ser dividida manualmente.

Para dividir uma região ativa específica, use o comando a seguir (executado em “hbase shell”).

```
hbase cli> split '<region name>'
```

Para dividir a tabela, use o comando a seguir (execute em “hbase shell”).

```
hbase cli> split 'hbase:storefile'
```

Considerações operacionais

HBaseos servidores regionais são usados BlockCache para armazenar leituras de dados na memória e BucketCache armazenar leituras de dados no disco local. Além disso, os servidores regionais usam MemStore para armazenar gravações de dados na memória e usam registros de gravação antecipada para armazenar gravações de dados HDFS antes que os dados sejam gravados no Amazon HBase StoreFiles S3. O desempenho de leitura do seu cluster estão relacionado à frequência com a qual um registro pode ser recuperado dos caches na memória ou no disco. Uma perda de cache faz com que o registro seja lido StoreFile no Amazon S3, que tem latência significativamente maior e maior desvio padrão do que a leitura. HDFS Além disso, as taxas máximas de solicitações para o Amazon S3 são menores do que as que podem ser obtidas no cache

local e, portanto, o armazenamento de dados no cache pode ser importante para workload com uso intensivo de leitura. Para obter mais informações sobre a performance do Amazon S3, consulte [Performance optimization](#), no Guia do usuário do Amazon Simple Storage Service.

Para melhorar o desempenho, recomendamos que você armazene em cache o máximo possível do seu conjunto de dados no armazenamento da EC2 instância. Como BucketCache usa o armazenamento de EC2 instâncias do servidor da região, você pode escolher um tipo de EC2 instância com um armazenamento de instâncias suficiente e adicionar EBS armazenamento da Amazon para acomodar o tamanho de cache necessário. Você também pode aumentar o BucketCache tamanho dos armazenamentos e EBS volumes de instâncias anexadas usando a `hbase.bucketcache.size` propriedade. A configuração padrão é 8.192 MB.

Para gravações, a frequência de MemStore descargas e o número de StoreFiles presentes durante compactações menores e maiores podem contribuir significativamente para um aumento nos tempos de resposta do servidor regional. Para um desempenho ideal, considere aumentar o tamanho do multiplicador de MemStore fluxo e HRegion bloco, o que aumenta o tempo decorrido entre as principais compactações, mas também aumenta o atraso na consistência se você usar uma réplica de leitura. Em alguns casos, você pode obter um melhor desempenho usando blocos de arquivos maiores (mas menos de 5 GB) para ativar a funcionalidade de upload de várias partes do Amazon S3. EMRFS O tamanho padrão EMR do bloco da Amazon é de 128 MB. Para obter mais informações, consulte [HDFSconfiguração](#). Raramente há clientes que excedem 1 GB de tamanho de bloco ao fazer a comparação do desempenho com liberações e compactações. Além disso, HBase as compactações e os servidores regionais funcionam de maneira ideal quando menos StoreFiles precisam ser compactados.

Tabelas podem demorar um tempo significativo para serem descartadas no Amazon S3, pois diretórios grandes precisam ser renomeados. Considere desabilitar tabelas em vez de as descartar.

Existe um processo de HBase limpeza que limpa WAL arquivos antigos e armazena arquivos. Com a EMR versão 5.17.0 e posterior da Amazon, o limpador é ativado globalmente e as seguintes propriedades de configuração podem ser usadas para controlar o comportamento do limpador.

Propriedade de configuração	Valor padrão	Descrição
<code>hbase.regionserver.hfilecleaner.larg e.thread.count</code>	1	O número de segmentos alocados para limpar expirou muito. HFiles

Propriedade de configuração	Valor padrão	Descrição
<code>hbase.regionserver.hfilecleaner.small.thread.count</code>	1	O número de segmentos alocados para limpar expirou pouco. HFiles
<code>hbase.cleaner.scan.dir.concurrent.size</code>	Defina como um quarto de todos os núcleos disponíveis.	O número de threads para verificar os oldWALS diretórios.
<code>hbase.oldwals.cleaner.thread.size</code>	2	O número de segmentos a serem limpos no WALs oldWALS diretório.

Com o Amazon EMR 5.17.0 e versões anteriores, a operação de limpeza pode afetar o desempenho da consulta ao executar cargas de trabalho pesadas, por isso recomendamos que você habilite o limpador somente fora dos horários de pico. O limpador tem os seguintes comandos de HBase shell:

- `cleaner_chore_enabled` consulta se o agente de limpeza está habilitado.
- `cleaner_chore_run` executa manualmente o agente de limpeza para remover arquivos.
- `cleaner_chore_switch` habilita ou desabilita o agente de limpeza e retorna ao seu estado anterior. Por exemplo, `cleaner_chore_switch true` habilita o agente de limpeza.

Propriedades para ajuste HBase de desempenho no Amazon S3

Os parâmetros a seguir podem ser ajustados para ajustar o desempenho da sua carga de trabalho quando você usa HBase no Amazon S3.

Propriedade de configuração	Valor padrão	Descrição
<code>hbase.bucketcache.size</code>	8,192	A quantidade de espaço em disco, em MB, reservada nos armazenamentos de EC2 instâncias da Amazon no servidor regional e EBS nos

Propriedade de configuração	Valor padrão	Descrição
		volumes para BucketCache armazenamento. A configuração se aplica a todas as instâncias do servidor de regiões. BucketCache Tamanhos maiores geralmente correspondem a um desempenho aprimorado
<code>hbase.hregion.memstore.flush.size</code>	134217728	O limite de dados, em bytes, no qual uma liberação de memstore para o Amazon S3 é acionada.
<code>hbase.hregion.memstore.block.multiplier</code>	4	Um multiplicador que determina o limite MemStore superior no qual as atualizações são bloqueadas. Se os MemStore excedentes forem <code>hbase.hregion.memstore.flush.size</code> multiplicados por esse valor, as atualizações serão bloqueadas. MemStore descargas e compactação podem ocorrer para desbloquear as atualizações.
<code>hbase.hstore.blockingStoreFiles</code>	10	O número máximo StoreFiles que pode existir em uma loja antes que as atualizações sejam bloqueadas.
<code>hbase.hregion.max.filesize</code>	10737418240	O tamanho máximo de uma região antes que ela seja dividida.

Desativar e restaurar um cluster sem perda de dados

Para desligar um EMR cluster da Amazon sem perder dados que não foram gravados no Amazon S3, você deve liberar seu MemStore cache para o Amazon S3 para gravar novos arquivos de armazenamento. Primeiro, é necessário desabilitar todas as tabelas. A seguinte configuração de etapas pode ser usada quando você adiciona uma etapa ao cluster. Para obter mais informações, consulte [Trabalhar com etapas usando o console AWS CLI e](#) no Amazon EMR Management Guide.

```
Name="Disable all tables",Jar="command-runner.jar",Args=["/bin/bash","/usr/lib/hbase/bin/disable_all_tables.sh"]
```

Como alternativa, você pode executar o seguinte comando bash diretamente.

```
bash /usr/lib/hbase/bin/disable_all_tables.sh
```

Depois de desativar todas as tabelas, limpe a hbase:meta tabela usando o HBase shell e o comando a seguir.

```
flush 'hbase:meta'
```

Em seguida, você pode executar um script de shell fornecido no EMR cluster da Amazon para MemStore limpar o cache. Você pode adicioná-lo como uma etapa ou executá-lo diretamente, usando a AWS CLI no cluster. O script desativa todas as HBase tabelas, o que faz com que o servidor MemStore de cada região seja transferido para o Amazon S3. Se o script for concluído com êxito, os dados persistirão no Amazon S3, e o cluster poderá ser terminado.

Para reiniciar um cluster com os mesmos HBase dados, especifique a mesma localização do Amazon S3 do cluster anterior na propriedade de configuração AWS Management Console ou usando a propriedade de hbase.rootdir configuração.

Registros de gravação antecipada (WAL) para Amazon EMR

Com o Amazon EMR 6.15 e versões posteriores, você pode gravar seus registros de HBase gravação antecipada do Apache () WAL na Amazon. EMR WAL Com EMR versões mais baixas da Amazon, quando você cria um cluster com a opção HBase no Amazon S3, WAL é o único HBase componente Apache que é armazenado no disco local para clusters, e você pode armazenar outros componentes, como o diretório raiz, armazenar arquivos (HFiles), metadados de tabelas e dados no Amazon S3.

Você pode usar EMR WAL a Amazon para recuperar dados que não foram transferidos para o Amazon S3. Para fazer backup completo de seus HBase clusters, opte por usar o EMR WAL serviço da Amazon. Nos bastidores, RegionServer grava seus registros HBase de gravação antecipada (WAL) no para a WAL Amazon. EMR

Caso seu cluster ou o AZ não estejam íntegros ou estejam indisponíveis, você pode criar um novo cluster, direcioná-lo para o mesmo diretório raiz do S3 e para o mesmo EMR WAL espaço de trabalho da Amazon e recuperar automaticamente os dados em alguns WAL minutos. Para obter mais informações, consulte [Restaurando da Amazon EMR WAL](#).

Note

A Amazon EMR retém seu registro de gravação antecipada e seus dados por 30 dias a partir do momento em que você cria seu cluster. Depois de 30 dias, a Amazon exclui EMR automaticamente sua Amazon EMR WAL e seus dados. No entanto, se você iniciar um novo cluster WAL habilitado a partir do mesmo diretório raiz do S3, poderá estender o uso do seu WAL por 30 dias a partir do momento da inicialização do novo cluster. A Amazon ainda EMR limpará todos WAL os dados do primeiro cluster após o período inicial de 30 dias. Para obter mais informações, consulte [Restaurando da Amazon EMR WAL](#).

As seções a seguir descrevem como configurar e usar a Amazon EMR WAL com seu EMR cluster HBase habilitado.

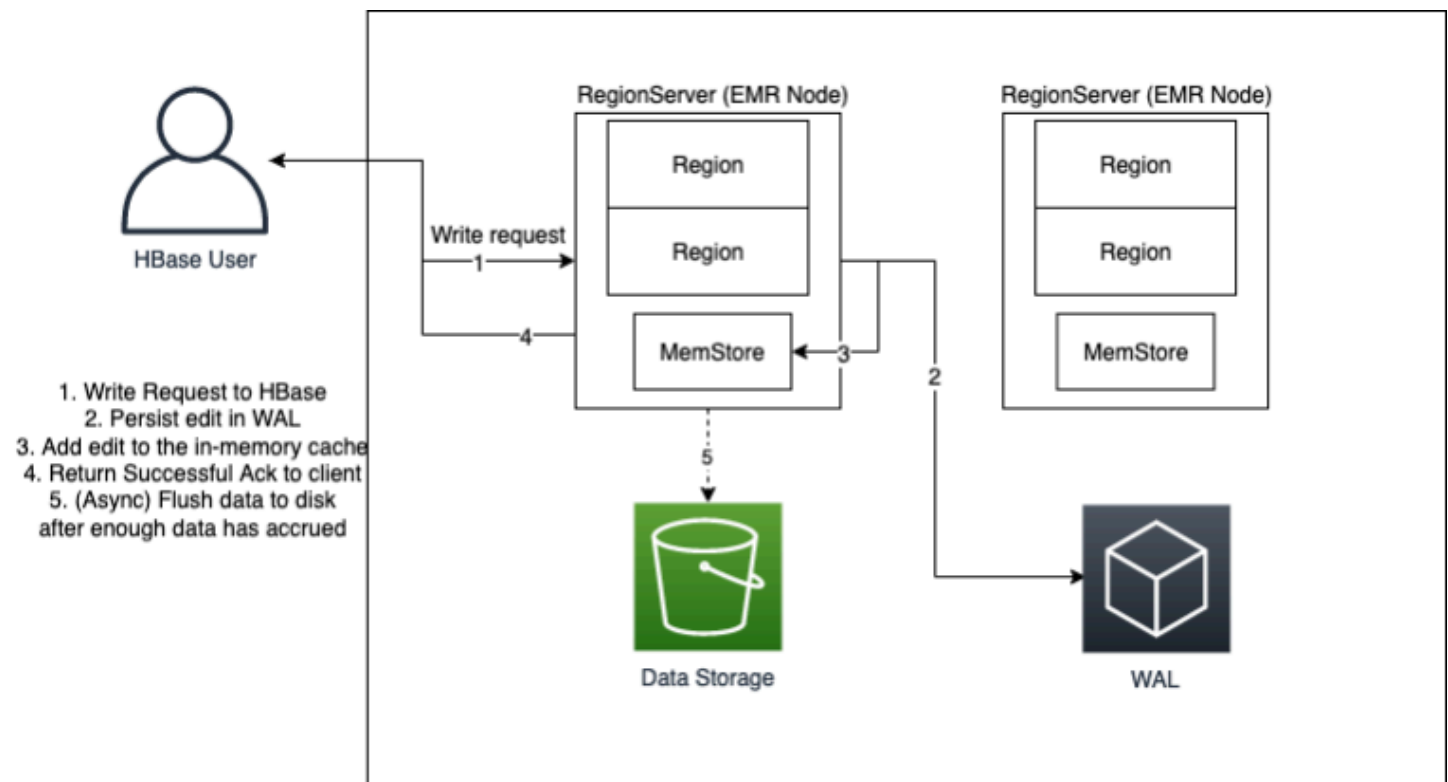
Tópicos

- [Espaços de EMR WAL trabalho da Amazon](#)
- [Permissões necessárias para a Amazon EMR WAL](#)
- [Habilitando a Amazon EMR WAL](#)
- [Restaurando da Amazon EMR WAL](#)
- [Usando configurações de segurança com a Amazon EMR WAL](#)
- [Acesse a Amazon EMR WAL por meio de AWS PrivateLink](#)
- [Entendendo os EMR WAL preços e as métricas da Amazon](#)
- [Marcação de espaços de trabalho WAL](#)
- [Considerações e regiões para a Amazon EMR WAL](#)
- [CLIReferência da Amazon EMR WAL \(EMRWAL\)](#)

Espaços de EMR WAL trabalho da Amazon

A Amazon EMR WAL adiciona o conceito de WAL espaços de trabalho. Um WAL espaço de trabalho é um contêiner lógico de. WALs Cada registro de gravação antecipada na Amazon EMR WAL é encapsulado por um espaço de trabalho. WAL Um EMR cluster grava WALs em exatamente um WAL espaço de trabalho que você configura na inicialização do cluster ou no, defaultWALworkspace se você não especificar um espaço de trabalho. WALos espaços de trabalho não estão relacionados a nenhuma HBase terminologia existente, como namespaces.

Você pode usar WAL espaços de trabalho para definir o escopo das EMR WAL IAM permissões da Amazon e incluir apenas os espaços de trabalho que o cluster precisa acessar. Você também pode marcar seu WAL espaço de trabalho para controle de acesso baseado em tags. Para obter mais informações sobre marcação, consulte [Marcação de espaços de trabalho WAL](#).



Permissões necessárias para a Amazon EMR WAL

Para que seu cluster se conecte à Amazon EMRWAL, o perfil da instância do cluster requer certas IAM permissões:

- A Amazon EMR WAL usa a função [AWSServiceRoleForEMRWAL](#) vinculada ao serviço para recuperar o status de um cluster. A Amazon cria EMR automaticamente essa função vinculada

ao serviço quando você cria um WAL espaço de trabalho, ou HBase cria a função vinculada ao serviço quando você configura um espaço de trabalho para a Amazon EMR WAL e a função vinculada ao serviço ainda não existe.

Antes de habilitar a Amazon EMR WAL para um cluster, você deve configurar as permissões para permitir a criação automática da função `AWSServiceRoleForEMRWAL` vinculada ao serviço. Para obter mais informações e um exemplo de declaração que adiciona esse recurso, consulte [Usando funções vinculadas ao serviço para registro antecipado](#).

- Como a Amazon EMR WAL usa o HBase Write Ahead Log (WAL), seus clusters devem usar HBaseWAL. A seguir estão as IAM permissões mínimas que você precisa para executar HBase. Adicione-os à política de permissões do seu perfil de instância:

```
emrwal:DeleteWal
emrwal:CreateWal
emrwal:CreateWorkspace
emrwal:AppendEdit
emrwal:ReplayEdits
emrwal:GetCurrentWalTime
emrwal:CompleteWalFlush
```

Note

Se você definir as permissões da Amazon apenas EMR WAL para o conjunto mínimo, alguns [EMRWALCLI](#) comandos não terão as permissões necessárias para serem executados.

Habilitando a Amazon EMR WAL

Use as etapas a seguir para habilitar a gravação na Amazon EMR WAL ao criar um cluster com AWS Command Line Interface o.

Note

Você não pode habilitar a Amazon EMR WAL para um cluster que já está em execução e não pode iniciar dois clusters com o mesmo diretório raiz do S3. Para obter mais informações, consulte [Considerações e regiões para a Amazon EMR WAL](#).

1. Antes de criar um cluster EMR WAL habilitado para Amazon, você deve adicionar as permissões necessárias ao perfil da instância que planeja usar com seu cluster. Para obter mais informações, consulte [Permissões necessárias para a Amazon EMR WAL](#).
2. Crie um cluster a partir do AWS CLI. Use a `--configurations` opção para fornecer um objeto de JSON configuração que especifique a `hbase.emr.wal.enabled` propriedade, conforme mostrado no exemplo abaixo.
 - Especifique o modo de armazenamento e a localização do diretório raiz no Amazon S3. A localização do Amazon S3 que você especificar deve estar na mesma região do seu EMR cluster, mas somente um cluster ativo pode usar o mesmo diretório HBase raiz no S3 por vez.
 - Crie seu cluster com a configuração de grupos de instâncias. Você não pode usar a Amazon EMR WAL com a configuração de frotas de instâncias. Para obter mais informações sobre a criação de clusters com grupos de instâncias, consulte [Configurar grupos de instâncias uniformes](#) no Amazon EMR Management Guide.
 - Para ver as etapas do console para criar um cluster e um `create-cluster` exemplo detalhado que usa o AWS CLI, consulte [Como criar um cluster com HBase](#).
3. WALPara habilitar o novo cluster, defina a `hbase.emr.wal.enabled` propriedade como `true`. O comando a seguir contém um JSON trecho com um exemplo de objeto de configuração.

```
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"  
    }  
  }  
]
```


Quando HBase estiver on-line no cluster recém-criado, HBase gravará WAL dados automaticamente na Amazon EMR WAL e usará a Amazon EMR WAL para fins de recuperação.

Example 1: Criação de um EMR cluster que usa a Amazon EMR WAL

```
[
  {
    "Classification": "hbase-site",
    "Properties": {
      "hbase.rootdir": "s3://MyBucket/MyHBaseStore"
    }
  },
  {
    "Classification": "hbase",
    "Properties": {
      "hbase.emr.storageMode": "s3",
      "hbase.emr.wal.enabled": "true"
    }
  }
]
```

Example 2: Criação de um EMR cluster com um espaço de WAL trabalho personalizado

```
[
  {
    "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"
    }
  }
]
```

Restaurando da Amazon EMR WAL

Como a Amazon EMR WAL do seu cluster original é retida por 30 dias, você pode restaurá-la e reutilizá-la WAL para um cluster recém-criado dentro desse período de 30 dias. Quando você executa um novo cluster a partir do mesmo diretório raiz do S3, o relógio de 30 dias é reiniciado a partir do horário de lançamento do novo cluster, desde que o período anterior de 30 dias não tenha expirado.

Use o procedimento a seguir para restaurar um cluster existente WAL com um novo. Esse processo pressupõe que você criou seu cluster original com a Amazon EMR WAL ativada.

1. Dentro de 30 dias após a criação WAL de um cluster habilitado, crie um novo cluster Região da AWS igual ao cluster original. O novo cluster pode estar na mesma AZ ou em uma AZ diferente na mesma região em que o cluster original foi criado.

Configure as propriedades do objeto para especificar o modo de armazenamento e a localização do diretório raiz no Amazon S3. A localização do Amazon S3 que você especificar deve estar na mesma região do seu EMR cluster, mas somente um cluster ativo pode usar o mesmo diretório HBase raiz no S3 por vez.

Para ver as etapas do console para criar um cluster e um `create-cluster` exemplo detalhado que usa o AWS CLI, consulte [Como criar um cluster com HBase](#).

2. Para usar a Amazon existente EMR WAL para o novo cluster, defina a `hbase.emr.wal.enabled` propriedade com `true`. O JSON trecho a seguir mostra um exemplo de objeto de configuração.

```
[
  {
    "Classification": "hbase-site",
    "Properties": {
      "hbase.rootdir": "s3://MyBucket/MyHBaseStore"
    }
  },
  {
    "Classification": "hbase",
    "Properties": {
      "hbase.emr.storageMode": "s3",
      "hbase.emr.wal.enabled": "true"
    }
  }
]
```

```
}  
]
```

Usando configurações de segurança com a Amazon EMR WAL

A Amazon criptografa EMR automaticamente os dados em trânsito entre seu cluster e o EMR WAL serviço da Amazon e os dados em repouso na Amazon EMR WAL. Para obter mais informações, consulte [Criptografia em repouso para a Amazon EMR WAL](#). Você também pode usar uma configuração de segurança para trazer suas próprias chaves do serviço AWS Key Management Service (KMS) e criptografar os dados que você armazena na Amazon EMR WAL.

Use um dos métodos a seguir para selecionar uma configuração de segurança ao criar um cluster:

Console

Em AWS Management Console, especifique a configuração em Configuração de segurança e par de EC2 chaves.

Security configuration and EC2 key pair - *optional* [Info](#)

Security configuration

Select your cluster encryption, authentication, authorization, and instance metadata service settings.



CLI

AWS CLI Em, defina o `--security-configuration` parâmetro ao usar o comando [create-cluster](#).

Para obter mais informações, consulte [Criptografia em repouso para a Amazon EMR WAL](#) e [Use configurações de segurança para configurar a segurança do cluster](#) no Amazon EMR Management Guide.

Para obter mais informações relacionadas à segurança WAL, consulte [Uso de funções vinculadas a serviços para registro antecipado](#).

Acesse a Amazon EMR WAL por meio de AWS PrivateLink

Se você quiser manter sua conexão dentro da AWS rede, a Amazon EMR WAL oferece AWS PrivateLink suporte. Para configurar AWS PrivateLink, use o AWS Management Console ou AWS

Command Line Interface (AWS CLI) para criar um VPC endpoint de interface que se conecta à Amazon EMR WAL. Para obter mais informações, consulte [Acessar um AWS serviço usando um VPC endpoint de interface](#) no AWS PrivateLink Guia.

As etapas básicas são:

1. Use o Amazon VPC Console para [criar um VPC endpoint](#). Selecione Endpoints e, em seguida, Create endpoint.
2. Mantenha a categoria Serviço como AWS serviços.
3. Na barra de pesquisa do painel Serviços **emrwal**, digite e selecione o serviço rotulado com `.amazonaws.region.emrwal.prod`.
4. Selecione seu VPC e salve o endpoint. Certifique-se de anexar ao VPC endpoint os mesmos grupos de segurança que você anexa ao EMR cluster.
5. Se quiser, agora você pode habilitar DNS nomes de host privados para seu novo endpoint. Defina Enable DNS hostnames e Enable DNS Support como `true` para seu VPC. Em seguida, selecione seu ID de endpoint, escolha Editar VPC configurações no menu Ações e habilite DNS nomes privados.
 - Os DNS nomes de host privados do endpoint seguirão o formato `prod.emrwal.region.amazonaws.com`
 - Se você não habilitar DNS nomes de host privados, a Amazon VPC fornecerá um nome de DNS endpoint para você no formato `endpointID.prod.emrwal.region.vpce.amazonaws.com`
6. Para usar seu AWS PrivateLink endpoint, modifique a `emr.wal.client.endpoint` configuração ao criar seu [cluster EMR WAL habilitado para Amazon](#), conforme mostrado no exemplo a seguir:

```
[
  {
    "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"
    }
  }
]

```

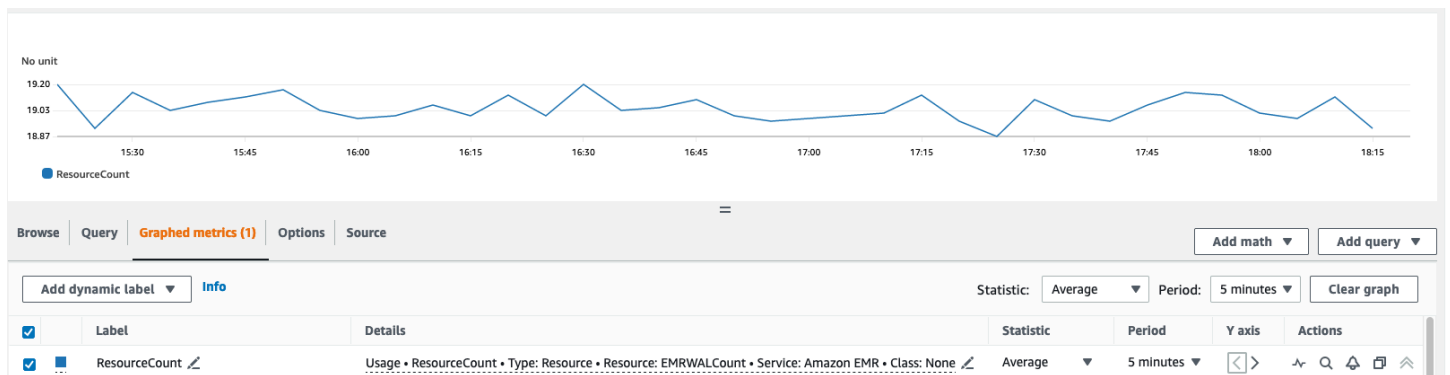
Você também pode usar a VPCE política para permitir ou restringir o acesso à Amazon EMR WAL APIs. Para obter mais informações, consulte [Controlar o acesso aos VPC endpoints usando políticas de endpoint](#) no AWS PrivateLink Guia.

Entendendo os EMR WAL preços e as métricas da Amazon

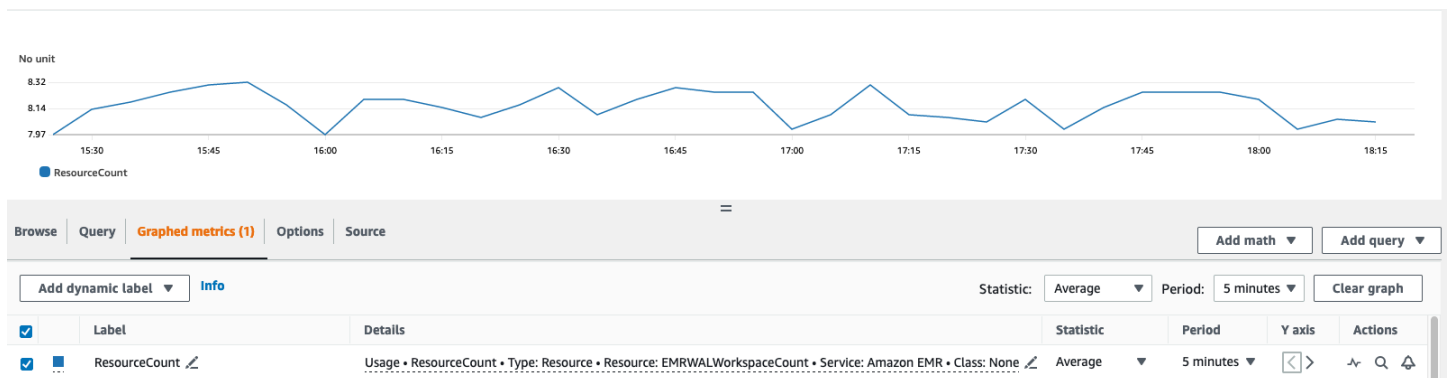
Unidade de cobrança de recursos principais	Detalhes
EMR-WAL-Read-GiB	API calls to read data from your table are billed as ReadRequestGiB. This includes Get and Scan operations. Reads are charged based on the sizes of the read items. Amazon EMR bills at a minimum of 1 byte. For example, if you read a 1234.12 bytes item, you're charged for 1235 bytes. Reads are aggregated every hour for billing and shown as GiBs.
EMR-WAL-Write-GiB	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.
EMR-WAL-WALHours	O número WALs que você armazena no serviço é cobrado como EMR-WAL-WALHours. EMR Amazon cria um WAL por HBase região. Por exemplo, se você criar 20 HBase tabelas, incluindo tabelas do sistema, e cada tabela tiver duas HBase regiões, você usará 28.800 WAL horas, calculadas como:

Unidade de cobrança de recursos principais	Detalhes
	<p>20 tables x 2 Regions per table x 1 WAL per Region x 30 days x 24 hours ----- 28,800 EMR-WAL-WALHours</p>

ExemploEMRWALCount:



ExemploEMRWALWorkspaceCount:



Marcação de espaços de trabalho WAL

Você pode adicionar tags a um espaço de trabalho ao criar um novo espaço de trabalho e pode adicionar, remover ou listar tags de um espaço de trabalho ativo para um cluster em execução. Você não pode marcar os recursos individuais na área de trabalho e não pode atualizar as tags existentes; em vez disso, remova as tags indesejadas da área de trabalho e substitua-as.

Você pode marcar espaços de trabalho a partir do EMRWAL CLI. Para obter uma lista de EMRWAL CLI comandos para marcar espaços de trabalho, consulte [CLIReferência da Amazon EMR WAL \(EMRWAL\)](#)

O exemplo de IAM política a seguir ilustra um cenário que permite CRUDL operações de espaço de trabalho somente com a chave `resource_tag_allow_test_key` e o valor de marcação adequados: `resource_tag_allow_test_value`

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Action": [
        "EMRWAL:*"
      ],
      "Effect": "Allow",
      "Resource": [
        "*",
        "*"
      ],
      "Condition": {
        "StringEquals": {
          "aws:ResourceTag/resource_tag_allow_test_key": [
            "resource_tag_allow_test_value"
          ]
        }
      }
    }
  ]
}
```

Para verificar se a tag agora é necessária para as operações do espaço de trabalho, use o [CLIReferência da Amazon EMR WAL \(EMRWAL\)](#) para chamar o `listTagsForResource` comando `tagAllowResourceTag` para o espaço de trabalho com a tag de recurso desejada. Se você configurou a condição corretamente, o comando será bem-sucedido.

```
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)
```

Considerações e regiões para a Amazon EMR WAL

Considerações sobre a Amazon EMR WAL

A lista a seguir descreve considerações e limitações importantes da Amazon EMRWAL:

- A Amazon EMR WAL está disponível para uso com as EMR versões 6.15.0 e superiores da Amazon.
- A Amazon EMR WAL é um serviço pago opcional. Você paga pelo que usa: leituras, gravações e armazenamento de dados. Para obter mais informações, consulte [Entendendo os EMR WAL preços e as métricas da Amazon](#) a página de [EMRpreços da Amazon](#).
- A Amazon EMR WAL usa o HBase Write Ahead Log (WAL). Para usar a Amazon EMRWAL, seus clusters devem usar HBaseWAL.
- Para habilitar a Amazon EMR WAL ao criar um cluster, você deve ter as permissões de função necessárias. Para obter mais informações, consulte [Uso de funções vinculadas a serviços para registro antecipado](#).
- Você deve habilitar a Amazon EMR WAL ao criar o cluster com o AWS Management Console,, ou AWS CLI API, e você deve usar a configuração de grupos de instâncias. Você não pode habilitar a Amazon EMR WAL em um cluster em execução se não tiver criado o cluster com a Amazon EMRWAL. Você também não pode editar as hbase-site configurações para habilitar a Amazon EMR WAL em um cluster em execução.
- Você só pode habilitar a Amazon EMR WAL em clusters que usam o Amazon S3 como diretório raiz.
- Você não pode ter vários clusters ativos no mesmo diretório HBase raiz no Amazon S3.
- Você não pode habilitar a Amazon EMR WAL em clusters de réplicas de leitura.
- WALé replicado em zonas de disponibilidade dentro do serviço gerenciado.
- WALsobrevive ao cluster e permanece disponível para o próximo cluster.
- Você não pode desativar EMR WAL a Amazon durante o lançamento ou quando seu cluster estiver operacional (em estado de execução).
- Para obter informações WAL e limites de espaço de trabalho, consulte [EMRendpoints e cotas da Amazon](#).

Disponibilidade regional para a Amazon EMR WAL

O EMR WAL serviço da Amazon está disponível no seguinte Regiões da AWS:

- `ap-northeast-1`: Ásia-Pacífico (Tóquio)
- `ap-southeast-1`: Ásia-Pacífico (Singapura)
- `ap-south-1`: Ásia-Pacífico (Mumbai)
- `ap-southeast-2`: Ásia-Pacífico (Sydney)
- `eu-central-1`: Europa (Frankfurt)
- `eu-north-1`: Europa (Estocolmo)
- `eu-west-1`: Europa (Irlanda)
- `us-east-1`: Leste dos EUA (Norte da Virgínia)
- `us-east-2`: Leste dos EUA (Ohio)
- `us-west-2`: Oeste dos EUA (Oregon)
- `ap-northeast-2`: Ásia-Pacífico (Seul)

CLIReferência da Amazon EMR WAL (EMRWAL)

A interface de linha de EMRWAL comando (EMRWAL CLI) é uma ferramenta unificada para gerenciar seu registro de gravação antecipada (WAL) para a Amazon. EMR É EMRWAL CLI fornecido com EMR clusters quando você ativa WAL no momento em que cria um cluster. Para obter mais informações sobre como habilitarWAL, consulte [Registros de gravação antecipada \(WAL\) para Amazon EMR](#).

EMRWAL CLIisso inclui os seguintes comandos:

Tópicos

- [createWorkspace](#)
- [deleteWal](#)
- [deleteWorkspace](#)
- [listTagsForResource](#)
- [listWals](#)
- [listWorkspaces](#)
- [tagResource](#)
- [untagResource](#)

createWorkspace

O createWorkspace comando cria um novo EMR WAL espaço de trabalho da Amazon.

Uso:

```
emrwal createWorkspace [-tags <tags>] [-e {endpoint}] [-r {Region}] -w {workspacename} [-h]
```

Exemplo:

```
aws emr createWorkspace -w examplews
```

deleteWal

O deleteWals comando exclui a Amazon EMR WAL que você especifica.

Uso:

```
emrwal deleteWal [-e {endpoint}] [-r {Region}] [-w {workspacename}] [-p <tablePrefix>] [-n <walName>] [-N <fullName>] [-R] [-m] [-h]
```

Exemplo:

```
emrwal deleteWal -w examplews -p hbasetable -n examplewal
```

deleteWorkspace

O deleteWorkspace comando exclui o EMR WAL espaço de trabalho da Amazon que você especificar.

Uso:

```
emrwal deleteWorkspace [-e {endpoint}] [-r {Region}] -w {workspacename} [-h]
```

Exemplo:

```
emrwal deleteWorkspace -w examplews
```

listTagsForResource

O `listTagsForResource` comando lista todas as tags de pares de valores-chave para o EMR WAL espaço de trabalho da Amazon que você especificar.

Uso:

```
emrwal listTagsForResource -arn {resource-arn} [-e {endpoint}] [-r {Region}] [-h]
```

Exemplo:

```
emrwal listTagsForResource -arn arn:aws:emrwal::1234567891234:workspace/examplews
```

listWals

O `listWals` comando lista toda a Amazon EMR WALs no espaço de trabalho que você especificar.

Uso:

```
emrwal listWals [-nextToken {token-string}] [-pageSize {integer}] [-e {endpoint}] [-r {Region}] [-w {workspacename}] [-p <tablePrefix>] [-M {integer}] [-h]
```

Exemplo:

```
emrwal listWals -w examplews
```

listWorkspaces

O `listWorkspaces` comando lista todos os EMR WAL espaços de trabalho da Amazon que estão disponíveis para você.

Uso:

```
emrwal listWorkspaces [-nextToken {token-string}] [-pageSize {integer}] [-e {endpoint}] [-r {Region}] [-M {integer}] [-h]
```

Exemplo:

```
emrwal listWorkspaces
```

tagResource

O `tagResource` comando atribui uma ou mais tags de pares de valores-chave ao EMR WAL espaço de trabalho da Amazon que você especificar.

Uso:

```
emrwal tagResource -arn {resource-arn} -tags <tags> [-e {endpoint}] [-r {Region}] [-h]
```

Exemplo:

```
emrwal tagResource -arn arn:aws:emrwal::1234567891234:workspace/examples -  
tags tag_key=tag_value
```

untagResource

O `untagResource` comando cancela a atribuição de uma ou mais tags de pares de valores-chave ao espaço de EMR WAL trabalho da Amazon que você especificar.

Uso:

```
emrwal untagResource -arn {resource-arn} -tagKeys <tagKeys> [-e {endpoint}] [-  
r {Region}] [-h]
```

Exemplo:

```
emrwal untagResource -arn arn:aws:emrwal::1234567891234:workspace/examples -  
tagKeys tag_key
```

Usando a HBase concha

Depois de criar um HBase cluster, a próxima etapa é conectar-se para que você HBase possa começar a ler e gravar dados (as gravações de dados não são suportadas em um cluster de réplica de leitura). Você pode usar o [HBaseshell](#) para testar comandos.

Para abrir a HBase concha

1. Use SSH para se conectar ao servidor principal no HBase cluster. Para obter informações sobre como se conectar ao nó EMR primário da Amazon usando SSH, consulte [Conectar ao nó primário usando SSH](#) o Amazon EMR Management Guide.
2. Executar `hbase shell`. O HBase shell é aberto com um prompt semelhante ao seguinte.

```
hbase(main):001:0>
```

Você pode emitir comandos de HBase shell a partir do prompt. Para obter mais informações sobre os comandos do shell e como chamá-los, digite `help` no HBase prompt e pressione Enter.

Criar uma tabela

O comando a seguir cria uma tabela chamada “t1”, que tem uma família de coluna exclusiva chamada de “f1”.

```
hbase(main):001:0>create 't1', 'f1'
```

Inserir um valor

O comando a seguir insere o valor “v1” para a linha “r1” na tabela “t1” e na coluna “f1”.

```
hbase(main):001:0>put 't1', 'r1', 'f1:col1', 'v1'
```

Obter um valor

O comando a seguir obtém os valores para a linha “r1” na tabela “t1”.

```
hbase(main):001:0>get 't1', 'r1'
```

Excluir uma tabela

O comando a seguir descarta e exclui a tabela “t1”.

```
hbase(main):001:0>drop 'ns1:t1', false
```

O valor booleano indica se você deseja ou não arquivar a tabela, então você pode defini-la como `true` se quiser salvá-la. Também é possível executar `drop 'ns1:t1'` sem booleano para arquivar a tabela.

Acesse HBase tabelas com o Hive

HBase e [Apache Hive](#) estão totalmente integrados, permitindo a execução massiva de cargas de trabalho de processamento paralelo diretamente nos dados armazenados em HBase. Para usar o Hive com HBase, geralmente você pode iniciá-los no mesmo cluster. No entanto, você pode iniciar o Hive e HBase em clusters separados. Executar HBase e o Hive separadamente em clusters diferentes pode melhorar o desempenho, pois isso permite que cada aplicativo use os recursos do cluster com mais eficiência.

Os procedimentos a seguir mostram como se conectar HBase em um cluster usando o Hive.

Note

Você só pode conectar um cluster do Hive a um único HBase cluster.

Para conectar o Hive ao HBase

1. Crie clusters separados com o Hive e HBase instalado ou crie um único cluster com ambos HBase e o Hive instalados.
2. Se você estiver usando clusters separados, modifique seus grupos de segurança para que HBase as portas do Hive estejam abertas entre esses dois nós primários.
3. Use SSH para se conectar ao nó primário do cluster com o Hive instalado. Para obter mais informações, consulte [Conecte-se ao nó primário usando SSH](#) o Amazon EMR Management Guide.
4. Execute o shell do Hive com o comando a seguir.

```
hive
```

5. (Opcional) Não é necessário fazer isso se HBase e o Hive estiver localizado no mesmo cluster. Conecte o HBase cliente em seu cluster Hive ao HBase cluster que contém seus dados. No exemplo a seguir, *public-DNS-name* é substituído pelo DNS nome público do nó primário do HBase cluster, por exemplo: `ec2-50-19-76-67.compute-1.amazonaws.com`.

```
set hbase.zookeeper.quorum=public-DNS-name;
```

6. Continue executando as consultas do Hive em seus HBase dados conforme desejado ou veja o próximo procedimento.

Para acessar HBase dados do Hive

- Depois que a conexão entre o Hive e os HBase clusters for estabelecida (conforme mostrado no procedimento anterior), você poderá acessar os dados armazenados no HBase cluster criando uma tabela externa no Hive.

O exemplo a seguir, quando executado a partir do prompt do Hive no nó primário, cria uma tabela externa que faz referência aos dados armazenados em uma HBase tabela chamada `inputTable`. Em seguida, você pode fazer referência `inputTable` nas instruções do Hive para consultar e modificar os dados armazenados no 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 ;
```

Para um caso de uso mais avançado e exemplos de combinação HBase com o Hive, consulte a postagem do blog sobre AWS Big Data, [Combine No, SQL e análises paralelas massivas usando o Apache HBase e o Apache Hive na Amazon](#). EMR

Usando HBase instantâneos

HBase usa uma funcionalidade de [captura instantânea](#) integrada para criar backups leves de tabelas. Em EMR clusters, esses backups podem ser exportados para o Amazon EMRFS S3 usando. Você pode criar um instantâneo no nó primário usando o HBase shell. Este tópico mostra como

executar esses comandos interativamente com o shell ou por meio de uma etapa usando `command-runner.jar` o AWS CLI ou AWS SDK for Java. Para obter mais informações sobre outros tipos de HBase backups, consulte [HBasebackup](#) na HBase documentação.

Criar um snapshot usando uma tabela

```
hbase snapshot create -n snapshotName -t tableName
```

Usando o `command-runner.jar` a partir do 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

Se o nome do snapshot não for exclusivo, a operação de criação falhará com um código de retorno -1 ou 255, mas talvez você não veja uma mensagem de erro afirmando o que deu errado. Para usar o mesmo nome de snapshot, exclua-o e recrie-lo em seguida.

Excluir um snapshot

```
hbase shell  
>> delete_snapshot 'snapshotName'
```

Exibir informações do snapshot

```
hbase snapshot info -snapshot snapshotName
```


Exportar um snapshot para o Amazon S3

Important

Se você não especificar um `-mappers` valor ao exportar um instantâneo, HBase usa um cálculo arbitrário para determinar o número de mapeadores. Esse valor pode ser muito grande, dependendo do tamanho da tabela, o que afeta negativamente os trabalhos em execução durante a exportação. Por esse motivo, recomendamos que você especifique o parâmetro `-mappers`, o parâmetro `-bandwidth` (que indica o consumo de largura de banda em megabytes por segundo) ou ambos para limitar os recursos de cluster usados pela operação de exportação. Como alternativa, você pode executar a operação de exportação de snapshot durante um período de baixa utilização.

```
hbase snapshot export -snapshot snapshotName \  
-copy-to s3://bucketName/folder -mappers 2
```

Usando `command-runner.jar` a partir do 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");
```

Importar um snapshot do Amazon S3

Embora seja uma importação, a HBase opção usada aqui ainda é `export`.

```
sudo -u hbase hbase snapshot export \  
-D hbase.rootdir=s3://bucketName/folder \  
-D hbase.snapshot.dir=s3://bucketName/folder \  
-D hbase.snapshot.name=snapshotName
```

```
-snapshot snapshotName \
-copy-to hdfs://masterPublicDNSName:8020/user/hbase \
-mappers 2
```

Usando `command-runner.jar` a partir do 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");
```

Restaurar uma tabela a partir de instantâneos dentro do shell HBase

```
hbase shell
>> disable tableName
>> restore_snapshot snapshotName
>> enable tableName
```

HBase atualmente não oferece suporte a todos os comandos de snapshot encontrados no HBase shell. Por exemplo, não há opção de HBase linha de comando para restaurar um snapshot, então você deve restaurá-lo dentro de um shell. Isso significa que `command-runner.jar` deve executar um comando Bash.

Note

Como o comando usado aqui é echo, é possível que seu comando shell ainda falhe mesmo que o comando executado pela Amazon EMR retorne um código de 0 saída. Verifique os logs de etapas se você optar por executar um comando shell como um etapa.

```
echo 'disable tableName; \
restore_snapshot snapshotName; \
enable tableName' | hbase shell
```

Veja a seguir a etapa usando a AWS CLI. Primeiro, crie o seguinte arquivo `snapshot.json`:

```
[
  {
    "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");
```

Configurar HBase

Embora as HBase configurações padrão devam funcionar para a maioria dos aplicativos, você pode modificá-las. HBase Para fazer isso, use as propriedades das classificações de HBase configuração. Para obter mais informações, consulte [Configurar aplicações](#).

O exemplo a seguir cria um cluster com um diretório HBase raiz alternativo baseado em um arquivo de configuração, `myConfig.json`, armazenado no Amazon S3.

Note

Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

```
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
```

O arquivo `myConfig.json` especifica a propriedade `hbase.rootdir` para a classificação de configuração `hbase-site`, conforme mostrado no exemplo a seguir. Substituir *ip-XXX-XX-XX-XXX.ec2.internal* com o DNS nome do host interno do nó primário do cluster.

```
[  
  {  
    "Classification": "hbase-site",  
    "Properties": {  
      "hbase.rootdir": "hdfs://ip-XXX-XX-XX-XXX.ec2.internal:8020/user/  
myCustomHBaseDir"  
    }  
  }  
]
```

Note

Com a Amazon EMR versão 5.21.0 e posterior, você pode substituir as configurações do cluster e especificar classificações de configuração adicionais para cada grupo de instâncias em um cluster em execução. Você faz isso usando o EMR console da Amazon, o AWS Command Line Interface (AWS CLI) ou AWS SDK o. Para obter mais informações, consulte [Supplying a Configuration for an Instance Group in a Running Cluster](#).

Alterações na alocação de memória em YARN

HBasenão está sendo executado como um YARN aplicativo, portanto, é necessário recalculer a memória alocada YARN e seus aplicativos, o que resulta em uma redução na memória geral

disponível se YARN estiver instalada. HBase Você deve levar isso em consideração ao planejar a co-localização de YARN aplicativos e HBase nos mesmos clusters. Os tipos de instância com menos de 64 GB de memória têm metade da memória disponível para NodeManager, que é então alocada para o HBase RegionServer. Por exemplo, tipos com memória maior que 64 GB, a HBase RegionServer memória é limitada a 32 GB. Como regra geral, YARN configurar a memória é um múltiplo da memória de tarefas do MapReduce redutor.

As tabelas em [Valores padrão para definições de configuração de tarefa](#) mostram as alterações YARN nas configurações com base na memória necessária para HBase.

HBases números de porta

Alguns números de porta escolhidos HBase são diferentes do padrão. A seguir estão as interfaces e portas para a HBase AmazonEMR.

HBase portas

Interface	Port (Porta)	Protocolo
HMaster	16000	TCP
HMasterUI	16010	HTTP
RegionServer	16020	TCP
RegionServer Informações	16030	HTTP
RESTservidor	8070	HTTP
RESTUI	8085	HTTP
Servidor Thrift	9090	TCP
Interface do usuário do servidor Thrift	9095	HTTP

⚠ Important

O `kms-http-port` é 9700 e o `kms-admin-port` é 9701 na EMR versão 4.6.0 e posterior da Amazon.

HBaseconfigurações do site para otimizar

Você pode definir qualquer uma ou todas as configurações do HBase site para otimizar o HBase cluster para a carga de trabalho do seu aplicativo. Recomendamos as seguintes configurações como ponto de partida na sua investigação.

`zookeeper.session.timeout`

O tempo limite padrão é de 40 segundos (40.000 ms). Se um servidor de regiões travar, este será o tempo necessário para o servidor mestre notar a ausência do servidor de regiões e iniciar a recuperação. Para ajudar o servidor mestre a se recuperar com mais rapidez, você pode reduzir esse valor para um período mais curto. O exemplo a seguir usa 30 segundos ou 30.000 ms:

```
[
  {
    "Classification": "hbase-site",
    "Properties": {
      "zookeeper.session.timeout": "30000"
    }
  }
]
```

`hbase.regionserver.handler.count`

Define o número de threads que o servidor de regiões mantém abertos para atender às solicitações de tabelas. O padrão de 10 é baixo, a fim de impedir que os usuários eliminem seus servidores de regiões ao usarem buffers de gravação grandes com um alto número de clientes simultâneos. A regra geral é manter esse número baixo quando a carga útil por solicitação se aproxima da faixa de MB (grandes entradas, digitalizações usando um cache grande) e alto quando a carga útil é pequena (obtenções, pequenas entradas, ICVs exclusões). O exemplo a seguir aumenta o número de threads abertos para 30:

```
[
```

```
[
  {
    "Classification": "hbase-site",
    "Properties": {
      "hbase.regionserver.handler.count": "30"
    }
  }
]
```

hbase.hregion.max.filesize

Esse parâmetro determina o tamanho, em bytes, das regiões individuais. Por padrão, ele é definido como 1073741824. Se você estiver gravando muitos dados em seu HBase cluster e isso estiver causando divisões frequentes, você pode aumentar esse tamanho para aumentar as regiões individuais. Isso reduz as divisões, mas exige mais tempo para fazer o balanceamento de carga de regiões de um servidor para outro.

```
[
  {
    "Classification": "hbase-site",
    "Properties": {
      "hbase.hregion.max.filesize": "1073741824"
    }
  }
]
```

hbase.hregion.memstore.flush.size

Esse parâmetro determina o tamanho máximo de memstore, em bytes, antes que ele seja liberado no disco. O padrão é 134217728. Se a sua workload é formada por curtos disparos contínuos de operações de gravação, convém aumentar esse limite para que todas as gravações permaneçam na memória durante o disparo contínuo e sejam liberadas no disco mais tarde. Isso pode aumentar o desempenho durante disparos contínuos.

```
[
  {
    "Classification": "hbase-site",
    "Properties": {
      "hbase.hregion.memstore.flush.size": "134217728"
    }
  }
]
```

]

Visualize a interface HBase do usuário

Note

A interface HBase do usuário usa HTTP conexões inseguras por padrão. Para habilitar secure HTTP (HTTPS), defina a `hbase.ssl.enabled` propriedade da `hbase-site` classificação `true` em sua [HBase configuração](#). Para obter mais informações sobre o uso de secure HTTP (HTTPS) para a interface do usuário HBase da web, consulte o [Guia de HBase referência do Apache](#).

HBase fornece uma interface de usuário baseada na web que você pode usar para monitorar seu HBase cluster. Quando você executa HBase na Amazon EMR, a interface web é executada no nó primário e pode ser visualizada usando o encaminhamento de portas, também conhecido como criação de um SSH túnel.

Para visualizar a interface HBase do usuário

1. Use SSH para criar um túnel até o nó primário e criar uma conexão segura. Para obter mais informações, consulte a [Opção 2, parte 1: Configurar um SSH túnel para o nó primário usando o encaminhamento dinâmico de portas](#) no Amazon EMR Management Guide.
2. Instale um navegador da Web com uma ferramenta de proxy, como o FoxyProxy plug-in do Firefox, para criar um SOCKS proxy para AWS domínios. Para obter mais informações, consulte a [Opção 2, parte 2: Definir configurações de proxy para visualizar sites hospedados no nó principal](#) no Amazon EMR Management Guide.
3. Com o proxy definido e a SSH conexão aberta, você pode visualizar a HBase interface do usuário abrindo uma janela do navegador com `http://master-public-dns-name:16010/master-status`, onde *master-public-dns-name* é o DNS endereço público do nó primário do cluster.



Home Table Details Local Logs Log Level Debug Dump Metrics Dump HBase Configuration

Master [redacted].ec2.internal

Region Servers

ServerName	Start time	Version	Requests Per Second	Num. P
[redacted].ec2.internal,16020,1461165084992	Wed Apr 20 15:11:24 UTC 2016	1.2.0	0	1
[redacted].ec2.internal,16020,1461165087881	Wed Apr 20 15:11:27 UTC 2016	1.2.0	0	2
Total:2			0	3

Você também pode ver HBase em Hue. Por exemplo, o seguinte mostra a tabela, t1, criada em [Usando a HBase concha](#):

The screenshot shows the Hue HBase Browser interface. The top navigation bar includes 'HUE', a home icon, and dropdown menus for 'Query Editors', 'Data Browsers', and 'Workflows'. The main header reads 'HBase Browser'. Below this, the breadcrumb path is 'Home - Bigtop / t1'. A search bar contains the query: 'row_key, row_prefix* +scan_len [col1, family:col2, fam3:, col_prefix* +3, fam:'. Below the search bar, two rows of data are displayed:

Row Key	Column Family	Value
r1	ft: col1	v1
r2	ft: col1	v2

Para obter mais informações sobre o Hue, consulte [Hue](#).

Exibir arquivos de log do HBase

Como parte de sua operação, HBase grava arquivos de log com detalhes sobre configurações, ações do daemon e exceções. Esses arquivos de log podem ser úteis para depurar problemas e HBase monitorar o desempenho.

Se você configurar o cluster para manter a persistência de arquivos de log para o Amazon S3, deve saber que os logs são gravados no Amazon S3 cada cinco minutos e, portanto, pode haver um pequeno atraso antes que os arquivos de log mais recentes estejam disponíveis.

Para visualizar HBase registros no nó primário

- Você pode visualizar os HBase registros atuais usando SSH para se conectar ao nó primário e navegar até o `/var/log/hbase` diretório. Esses logs não estão disponíveis depois que o cluster é encerrado, a menos que você habilite o registro em log no Amazon S3 quando o cluster for iniciado.

Para visualizar HBase registros no Amazon S3

- Para acessar HBase logs e outros logs de cluster no Amazon S3 e disponibilizá-los após o encerramento do cluster, especifique um bucket do Amazon S3 para receber esses registros ao criar o cluster. Isso é feito usando a opção `--log-uri`. Para obter mais informações sobre como habilitar o registro em seu cluster, consulte [Configurar registro e depuração \(opcional\)](#) no Amazon EMR Management Guide.

Monitor HBase com Ganglia

O projeto de código-fonte aberto Ganglia é um sistema distribuído e escalável, projetado para monitorar clusters e grades e, ao mesmo tempo, minimizar o impacto sobre o seu desempenho. Quando você habilita o Ganglia no seu cluster, pode gerar relatórios e visualizar o desempenho do cluster como um todo, bem como inspecionar o desempenho de instâncias de nós individuais. Para obter mais informações sobre o projeto de código-fonte aberto Ganglia, consulte <http://ganglia.info/>. Para obter mais informações sobre o uso do Ganglia com EMR clusters da Amazon, consulte.

[Ganglia](#)

Depois que o cluster for lançado com o Ganglia configurado, você poderá acessar os gráficos e relatórios do Ganglia usando a interface gráfica em execução no nó primário.

O Ganglia armazena arquivos de log no nó primário do `/mnt/var/lib/ganglia/rrds/` diretório. Versões de lançamento anteriores da Amazon EMR podem armazenar arquivos de log no `/var/log/ganglia/rrds/` diretório.

Para configurar um cluster para o Ganglia e usar o HBase AWS CLI

- Use um comando `create-cluster` semelhante ao seguinte:

```
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

Se a função de EMR serviço padrão da Amazon e o perfil de EC2 instância da Amazon não existirem, ocorrerá um erro. Use o comando `aws emr create-default-roles` para criá-los e, em seguida, tente novamente.

Para obter mais informações, consulte [EMRos comandos da Amazon no AWS CLI](#).

Para visualizar HBase métricas na interface web do Ganglia

1. Use SSH para criar um túnel até o nó primário e criar uma conexão segura. Para obter mais informações, consulte a [Opção 2, parte 1: Configurar um SSH túnel para o nó primário usando o encaminhamento dinâmico de portas](#) no Amazon EMR Management Guide.
2. Instale um navegador da Web com uma ferramenta de proxy, como o FoxyProxy plug-in do Firefox, para criar um SOCKS proxy para AWS domínios. Para obter mais informações, consulte a [Opção 2, parte 2: Definir configurações de proxy para visualizar sites hospedados no nó principal](#) no Amazon EMR Management Guide.
3. Com o proxy definido e a SSH conexão aberta, você pode visualizar as métricas do Ganglia abrindo uma janela do navegador com `http://master-public-dns-name/ganglia/`, onde *master-public-dns-name* é o DNS endereço público do servidor mestre no HBase cluster.

Para visualizar os arquivos de log do Ganglia no nó primário

- Se o cluster ainda estiver em execução, você poderá acessar os arquivos de log usando SSH para se conectar ao nó primário e navegar até o `/mnt/var/lib/ganglia/rrds/` diretório. Para EMR 3.x, navegue até o `/var/log/ganglia/rrds` diretório. Para obter mais informações, consulte [Conecte-se ao nó primário usando SSH](#) o Amazon EMR Management Guide.

Visualizar arquivos de log do Ganglia no Amazon S3

- Os arquivos de log do Ganglia não são gravados automaticamente no Amazon S3 mesmo se você habilitar o registro em log para seu cluster. Para visualizar os arquivos de log do Ganglia no Amazon S3, envie manualmente os logs do `/mnt/var/lib/ganglia/rrds/` para o bucket do S3.

Migrando de versões anteriores HBase

Para migrar dados de uma HBase versão anterior, consulte [Atualização, número da HBase versão e compatibilidade](#) no Guia de referência do ApacheHBase. Talvez seja necessário prestar atenção especial aos requisitos de atualização das versões anteriores à 1.0 do HBase

HBase histórico de lançamentos

A tabela a seguir lista a versão HBase incluída em cada versão de lançamento da AmazonEMR, junto com os componentes instalados com o aplicativo. Para obter as versões dos componentes em cada versão, consulte a seção Versão de componentes da sua versão em [Versões de lançamento do Amazon EMR 7.x](#), [Versões de lançamento do Amazon EMR 6.x](#) ou [Versões de lançamento do Amazon EMR 5.x](#).

HBase informações sobre a versão

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com HBase
		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
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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com HBase
emr-7.1.0	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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.14.2	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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.12.2	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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.11.3	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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server
emr-5.8.3	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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-resourcemanager, 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-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-region-server, 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-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, 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-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-region-server, 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-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-region-server, 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-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-resourcemanager, 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-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-region-server, 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-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, 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-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, 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-resourcemanager, hbase-hmaster, hbase-client, hbase-regionserver, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HBaseVersão	Componentes instalados com 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-httfs-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, 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-httfs-server, hadoop-mapred, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hbase-hmaster, hbase-client, hbase-region-server, hbase-rest-server, hbase-thrift-server, zookeeper-client, zookeeper-server

Apache HCatalog

HCatalog é uma ferramenta que permite acessar tabelas de metastore do Hive no Pig, no Spark SQL e/ou em aplicativos personalizados. MapReduce HCatalog tem uma REST interface e um cliente de linha de comando que permitem criar tabelas ou realizar outras operações. Em seguida, você escreve seus aplicativos para acessar as tabelas usando HCatalog bibliotecas. Para ter mais informações, consulte [Usar o HCatalog](#). HCatalog está incluído na EMR versão 4.4.0 e posterior da Amazon.

HCatalog na Amazon, a EMR versão 5.8.0 e posterior suporta o uso do AWS Glue Data Catalog como metastore para o Hive. Para obter mais informações, consulte [Usando o AWS Glue Data Catalog como metastore para o Hive](#).

A tabela a seguir lista a versão HCatalog incluída na versão mais recente da série EMR 6.x da Amazon, junto com os componentes com os quais a Amazon é EMR instalada. HCatalog

Para a versão dos componentes instalados HCatalog nesta versão, consulte [Versões de componentes da versão 6.15.0](#).

HCatalog informações da versão do emr-6.15.0

Gravadora EMR de lançamento da Amazon	HCatalog Versão	Componentes instalados com o 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

A tabela a seguir lista a versão HCatalog incluída na versão mais recente da série Amazon EMR 5.x, junto com os componentes com os quais a Amazon é EMR instalada. HCatalog

Para a versão dos componentes instalados HCatalog nesta versão, consulte Versões de componentes da [versão 5.36.2](#).

HCatalog informações da versão do emr-5.36.2

Gravadora EMR de lançamento da Amazon	HCatalog Versão	Componentes instalados com o 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

Tópicos

- [Criar um cluster com o HCatalog](#)
- [Usando HCatalog](#)
- [Exemplo: Crie uma HCatalog tabela e grave nela usando o Pig](#)
- [HCatalog histórico de lançamentos](#)

Criar um cluster com o HCatalog

Embora HCatalog esteja incluído no projeto Hive, você deve instalá-lo como seu próprio aplicativo.

Para iniciar um cluster com HCatalog instalado usando o console

O procedimento a seguir cria um cluster com HCatalog instalado. Para obter mais informações sobre a criação de clusters usando o console, incluindo opções avançadas, consulte [Planejar e configurar clusters](#) no Amazon EMR Management Guide.

1. Abra o EMR console da Amazon em <https://console.aws.amazon.com/emr>.
2. Escolha Create cluster (Criar cluster) para usar Quick Create.
3. Para o campo Software Configuration (Configuração do software), selecione Amazon versão emr-4.4.0 ou posterior.
4. No campo Selecionar aplicativos, escolha Todos os aplicativos ou HCatalog.
5. Selecione outras opções conforme necessário e escolha Create cluster (Criar cluster).

Para iniciar um cluster HCatalog usando o AWS CLI

- Crie o cluster usando o seguinte comando:

Note

Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

```
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
```

Usando HCatalog

Você pode usar HCatalog em vários aplicativos que usam o metastore Hive. Os exemplos nesta seção mostram como criar uma tabela e usá-la no contexto do Pig and SQL Spark.

Desative a gravação direta ao usar HCatalog HStorer

Sempre que um aplicativo usa [HCatStorer](#) para gravar em uma HCatalog tabela armazenada no Amazon S3, desative o recurso de gravação direta da Amazon. EMR Por exemplo,

desative a gravação direta ao usar o STORE comando Pig ou ao executar trabalhos do Sqoop que gravam HCatalog tabelas no Amazon S3. Você pode desativar o recurso de gravação direta definindo as configurações `mapred.output.direct.NativeS3FileSystem` e `mapred.output.direct.EmrFileSystem` como `false`. O exemplo a seguir demonstra como definir essas configurações usando o Java.

```
Configuration conf = new Configuration();
conf.set("mapred.output.direct.NativeS3FileSystem", "false");
conf.set("mapred.output.direct.EmrFileSystem", "false");
```

Crie uma tabela usando o HCat CLI e use esses dados no Pig

Crie o seguinte script, `impressions.q`, no seu 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');
```

Execute o script usando HCatCLI:

```
% 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
```

Abra o shell do Grunt e acesse os dados em `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,Irpv3oiu0I5QNQiwSSTIshrLdo9cM1,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,q2Awfnpe0JAvhInaIp0VGx9Kts0oP0,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)

```

Acessando a tabela usando o Spark SQL

Este exemplo cria um Spark a DataFrame partir da tabela criada no primeiro exemplo e mostra as primeiras 20 linhas:

```

% 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|q2Awfnp0JAvhInaI...|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|cjbTpruoiEtqLuMX...|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|pTne9k62k114QViXI...|RVxJVIQousjxUVI3r...| pixnet.net|
Mozilla/5.0 (Maci...|1bG0KiBD2xmui90kF...| 33.176.101.80|2009-04-13-08-05|
+-----+-----+-----+-----+
+-----+-----+-----+-----+

```

```
only showing top 20 rows
```

```
scala>
```

Exemplo: Crie uma HCatalog tabela e grave nela usando o Pig

Você pode criar uma HCatalog tabela e usar o Apache Pig para gravar nela HCatStorer usando uma fonte de dados no Amazon S3. HCatalog exige que você desative a gravação direta ou a operação falhará silenciosamente. Defina as configurações `mapred.output.direct.NativeS3FileSystem` e `mapred.output.direct.EmrFileSystem` como `false` usando a classificação `mapred-site` ou manualmente no shell do Grunt. O exemplo a seguir mostra uma tabela criada usando o HCatCLI, seguida por comandos executados no shell Grunt para preencher a tabela a partir de um arquivo de dados de amostra no Amazon S3.

Para executar este exemplo, [conecte-se ao nó principal usando SSH](#).

Crie um arquivo de HCatalog script, `wikicount.q`, com o conteúdo a seguir, que cria uma HCatalog tabela chamada `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';
```

Use um HCat CLI comando para executar o script a partir do arquivo.

```
hcat -f wikicount.q
```

Em seguida, inicie o shell do Grunt com a opção `-useHCatalog`, defina as configurações para desativar a gravação direta, carregue os dados de um local do S3 e, em seguida, grave os resultados na tabela `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();

```

HCatalog histórico de lançamentos

A tabela a seguir lista a versão HCatalog incluída em cada versão de lançamento da AmazonEMR, junto com os componentes instalados com o aplicativo. Para obter as versões dos componentes em cada versão, consulte a seção Versão de componentes da sua versão em [Versões de lançamento do Amazon EMR 7.x](#), [Versões de lançamento do Amazon EMR 6.x](#) ou [Versões de lançamento do Amazon EMR 5.x](#).

HCatalog informações sobre a versão

Gravadora Amazon EMR Release	HCatalog Versão	Componentes instalados com 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-server, 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-librar

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com HCatalog
		y, 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com HCatalog
emr-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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	HCatalogVersão	Componentes instalados com 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 é um data warehouse e um pacote de análises de código aberto executado além de um cluster do Hadoop. Os scripts do Hive usam uma linguagem SQL semelhante chamada Hive QL (linguagem de consulta) que abstrai os modelos de programação e oferece suporte às interações típicas do data warehouse. O Hive permite que você evite as complexidades de escrever trabalhos Tez com base em gráficos acíclicos direcionados (DAGs) ou MapReduce programas em uma linguagem de computador de nível inferior, como Java.

O Hive amplia o SQL paradigma ao incluir formatos de serialização. Você também pode personalizar o processamento de consultas com a criação de um esquema de tabela que corresponda a seus dados, sem tocar nos próprios dados. Embora SQL só suportem tipos de valores primitivos (como datas, números e cadeias de caracteres), os valores da tabela do Hive são elementos estruturados, como JSON objetos, qualquer tipo de dados definido pelo usuário ou qualquer função escrita em Java.

Para obter mais informações sobre o Hive, consulte <http://hive.apache.org/>.

A tabela a seguir lista a versão do Hive incluída na versão mais recente da série Amazon EMR 7.x, junto com os componentes que a Amazon EMR instala com o Hive.

Para a versão dos componentes instalados com o Hive nesta versão, consulte [Versões de componentes da versão 7.2.0](#).

Informações sobre a versão do Hive para o emr-7.2.0

Gravadora EMR de lançamento da Amazon	Versão do Hive	Componentes instalados com o 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

Gravadora EMR de lançamento da Amazon	Versão do Hive	Componentes instalados com o 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

A tabela a seguir lista a versão do Hive incluída na versão mais recente da série Amazon EMR 6.x, junto com os componentes que a Amazon EMR instala com o Hive.

Para obter a versão dos componentes instalados com o Hive nessa versão, consulte [Release 6.15.0 Component Versions](#).

Informações de versão do Hive para o emr-6.15.0

Gravadora EMR de lançamento da Amazon	Versão do Hive	Componentes instalados com o 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

A tabela a seguir lista a versão do Hive incluída na versão mais recente da série Amazon EMR 5.x, junto com os componentes que a Amazon EMR instala com o Hive.

Para a versão dos componentes instalados com o Hive nesta versão, consulte [Versões de componentes da versão 5.36.2](#).

Informações sobre a versão do Hive para o emr-5.36.2

Gravadora EMR de lançamento da Amazon	Versão do Hive	Componentes instalados com o 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

A partir do Amazon EMR 5.18.0, você pode usar o repositório de EMR artefatos da Amazon para criar seu código de trabalho com base nas versões exatas das bibliotecas e dependências que estão disponíveis em versões específicas da Amazon. Para ter mais informações, consulte [Verificar dependências usando o repositório de artefatos do Amazon EMR](#).

Tópicos

- [Diferenças e considerações sobre o Hive na Amazon EMR](#)
- [Configurar um metastore externo para o Hive](#)
- [Usar o driver JDBC do Hive](#)
- [Melhorar a performance do Hive](#)
- [Usando o Hive Live Long and Process \(\) LLAP](#)

- [Criptografia no Hive](#)
- [Histórico de versões do Hive](#)

Diferenças e considerações sobre o Hive na Amazon EMR

Diferenças entre o Apache Hive na Amazon EMR e o Apache Hive

[Esta seção descreve as diferenças entre o Hive na Amazon EMR e as versões padrão do Hive disponíveis em <http://svn.apache.org/viewvc/hive/branches/>.](#)

Autorização do Hive

A Amazon EMR suporta a [autorização do Hive](#) paraHDFS, mas não paraEMRFS, o Amazon S3. EMROs clusters da Amazon são executados com a autorização desativada por padrão.

Comportamento de mesclagem de arquivos do Hive com o Amazon S3

O Apache Hive mesclará arquivos pequenos no final de um trabalho somente mapa se `hive.merge.mapfiles` for "true", e a mesclagem apenas será acionada se o tamanho médio de saída do trabalho for menor do que a configuração `hive.merge.smallfiles.avgsize`. O Amazon EMR Hive tem exatamente o mesmo comportamento se o caminho de saída final estiver ativado. HDFS Se o caminho de saída estiver no Amazon S3, o parâmetro `hive.merge.smallfiles.avgsize` será ignorado. Nessa situação, a tarefa de mesclagem sempre será acionada se `hive.merge.mapfiles` estiver definido como true.

ACIDtransações e Amazon S3

O Amazon EMR 6.1.0 e versões posteriores oferecem suporte às transações do Hive ACID (atomicidade, consistência, isolamento, durabilidade) para que esteja em conformidade com as ACID propriedades de um banco de dados. Com esse recurso, você pode executarINSERT,, UPDATEDELETE, e MERGE operações em tabelas gerenciadas do Hive com dados no Amazon Simple Storage Service (Amazon S3).

Hive Live Long and Process () LLAP

[LLAPa funcionalidade](#) adicionada na versão 2.0 do Apache Hive padrão não é suportada no Hive 2.1.0 na Amazon EMR versão 5.0.

A EMR versão 6.0.0 e posterior da Amazon oferece suporte à funcionalidade Live Long and Process (LLAP) para o Hive. Para obter mais informações, consulte [Usando o Hive. LLAP](#)

Diferenças no Hive entre as EMR versões 4.x e 5.x da Amazon

Esta seção aborda as diferenças a serem consideradas antes de migrar uma implementação do Hive do Hive versão 1.0.0 na Amazon versão 4.x para o Hive 2.x na EMR Amazon versão 5.x. EMR

Considerações e diferenças operacionais

- Support adicionado para [transações ACID \(atomicidade, consistência, isolamento e durabilidade\)](#): essa diferença entre o Hive 1.0.0 no Amazon EMR 4.x e o Apache Hive padrão foi eliminada.
- As gravações diretas no Amazon S3 foram eliminadas: essa diferença entre o Hive 1.0.0 na EMR Amazon e o Apache Hive padrão foi eliminada. O Hive 2.1.0 na Amazon EMR versão 5.x agora cria, lê e grava em arquivos temporários armazenados no Amazon S3. Como resultado, para ler e gravar na mesma tabela, você não precisa mais criar uma tabela temporária no sistema de HDFS arquivos local do cluster como solução alternativa. Se você usar buckets com versionamento, certifique-se de gerenciar esses arquivos temporários conforme descrito abaixo.
- Gerenciar arquivos temporários ao serem usados buckets do Amazon S3 com versionamento: quando você executa consultas do Hive em que o destino dos dados gerados é o Amazon S3, muitos arquivos e diretórios e temporários são criados. Este é o novo comportamento, conforme descrito anteriormente. Se você usar buckets do S3 com versionamento, esses arquivos temporários sobrecarregarão o Amazon S3 e resultarão em custos caso não sejam excluídos. Ajuste suas regras de ciclo de vida para que os dados com um prefixo `/_tmp` sejam excluídos após um curto período, como cinco dias. Consulte [Especificar uma configuração de ciclo de vida](#) para obter mais informações.
- Log4j atualizado para log4j 2: se você usa log4j, talvez precise alterar sua configuração de registro em log devido a este upgrade. Para obter detalhes, consulte [Apache log4j 2](#).

Diferenças e considerações de desempenho

- Diferenças de desempenho com o Tez: com a EMR versão 5.x da Amazon, o Tez é o mecanismo de execução padrão para o Hive, em vez de. MapReduce O Tez fornece melhor desempenho para a maioria dos fluxos de trabalho.
- Tabelas com muitas partições: consultas que geram um grande número de partições dinâmicas podem falhar e consultas que selecionam tabelas com muitas partições podem demorar mais do

que o esperado para serem executadas. Por exemplo, um seleção de 100.000 partições pode demorar 10 minutos ou mais.

Recursos adicionais do Hive na Amazon EMR

A Amazon EMR amplia o Hive com novos recursos que oferecem suporte à integração do Hive com outros AWS serviços, como a capacidade de ler e gravar no Amazon Simple Storage Service (Amazon S3) e no DynamoDB.

Variáveis no Hive

Você pode incluir variáveis nos seus scripts usando o sinal de cifrão e chaves.

```
add jar ${LIB}/jsonserde.jar
```

Você transmite os valores dessas variáveis ao Hive na linha de comando usando o parâmetro `-d`, como no exemplo a seguir:

```
-d LIB=s3://elasticmapreduce/samples/hive-ads/lib
```

Você também pode transmitir os valores em etapas que executam scripts Hive.

Para transmitir valores variáveis em etapas do Hive usando o console

1. Abra o EMR console da Amazon em <https://console.aws.amazon.com/emr>.
2. Selecione Criar cluster.
3. Na seção Steps (Etapas), para Add Step (Adicionar etapa), escolha Hive Program (Programa Hive) na lista e depois Configure and add (Configurar e adicionar).
4. Na caixa de diálogo Add Step (Adicionar etapa), especifique os parâmetros usando a tabela a seguir como guia e escolha Add (Adicionar).

Campo	Ação
Localização do script no S3*	Especifique URI onde seu script reside no Amazon S3. O valor deve estar no formato <i>BucketName /path/ScriptName</i> . Por exemplo: <code>s3://elasticmapreduce/samples/hive-ads/libs/response-time-stats.q</code> .
Localização do S3 de entrada	Opcionalmente, especifique URI onde seus arquivos de entrada residem no Amazon S3. O valor deve estar no formato <i>BucketName /path/</i> . Se especificado, isso será passado para o script do Hive como um parâmetro chamado <code>INPUT</code> . Por exemplo: <code>s3://elasticmapreduce/samples/hive-ads/tables/</code> .
Local do S3 de saída	Opcionalmente, especifique URI onde você deseja que a saída seja armazenada no Amazon S3. O valor deve estar no formato <i>BucketName /path</i> . Se especificado, isso será passado para o script do Hive como um parâmetro chamado <code>OUTPUT</code> . Por exemplo: <code>s3://mybucket/hive-ads/output/</code> .
Argumentos	<p>Opcionalmente, insira uma lista de argumentos (strings separadas por espaços) para transmitir ao Hive. Se você definiu uma variável de caminho em seu script do Hive chamada <code> \${SAMPLE}</code>, por exemplo:</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>Para transmitir um valor para a variável, digite o seguinte na janela Arguments (Argumentos):</p> <pre>-d SAMPLE=s3://elasticmapreduce/samples/hive-ads/</pre>

Campo	Ação
Ação em caso de falha	<p>Isso determina o que faz o cluster em resposta a qualquer erro. Os valores possíveis para essa configuração são:</p> <ul style="list-style-type: none"> • Terminate cluster (Encerrar o cluster): se a etapa falhar, encerrar o cluster. Se o cluster tiver a proteção de encerramento ativada, o AND keep alive habilitado, ele não será encerrado. • Cancel and wait (Cancelar e aguardar): se a etapa falhar, cancele as etapas restantes. Se o cluster tiver o keep alive habilitado, ele não será encerrado. • Continue (Continuar): se a etapa falhar, continue na próxima etapa.

5. Selecione os valores conforme necessário e escolha Create cluster (Criar cluster).

Para passar valores variáveis para as etapas do Hive usando o AWS CLI

Para passar valores variáveis para as etapas do Hive usando o AWS CLI, use o `--steps` parâmetro e inclua uma lista de argumentos.

Note

Os caracteres de continuação de linha do Linux (`\`) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (`^`).

```
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/]
```

Para obter mais informações sobre o uso dos EMR comandos da Amazon no AWS CLI, consulte <https://docs.aws.amazon.com/cli/latest/reference/emr>.

Para passar valores variáveis para as etapas do Hive usando o Java SDK

- O exemplo a seguir demonstra como passar variáveis em etapas usando o SDK. Para obter mais informações, consulte [Classe StepFactory](#) na AWS SDK for Java API referência.

```
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")));
```

Consultas do Amazon EMR Hive para acomodar esquemas parciais do DynamoDB

O Amazon EMR Hive fornece flexibilidade máxima ao consultar tabelas do DynamoDB, permitindo que você especifique um subconjunto de colunas nas quais você pode filtrar dados, em vez de exigir que sua consulta inclua todas as colunas. Essa técnica de consulta de esquema parcial é eficaz quando você tem uma esquema de banco de dados esparso e deseja filtrar registros com base em algumas colunas, por exemplo, em carimbos de data/hora.

O exemplo a seguir mostra como usar uma consulta do Hive para:

- Crie uma tabela do DynamoDB.
- Selecione um subconjunto de itens (linhas) no DynamoDB e restrinja ainda mais os dados para determinadas colunas.
- Copie os dados resultantes para o 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";
```

A tabela a seguir mostra a sintaxe de consulta para a seleção de qualquer combinação de itens do DynamoDB.

Exemplo de consulta	Descrição do resultado
<code>SELECT * FROM <i>table_name</i> ;</code>	Seleciona todos os itens (linhas) de uma determinada tabela e inclui dados de todas as colunas disponíveis para esses itens.
<code>SELECT * FROM <i>table_name</i> WHERE <i>field_name</i> =<i>value</i>;</code>	Seleciona alguns itens (linhas) de uma determinada tabela e inclui dados de todas as colunas disponíveis para esses itens.
<code>SELECT <i>column1_name</i> , <i>column2_name</i> , <i>column3_name</i> FROM <i>table_name</i> ;</code>	Seleciona todos os itens (linhas) de uma determinada tabela e inclui dados de algumas colunas disponíveis para esses itens.
<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>	Seleciona alguns itens (linhas) de uma determinada tabela e inclui dados de algumas colunas disponíveis para esses itens.

Copiar dados entre tabelas do DynamoDB em diferentes regiões da AWS

O Amazon EMR Hive fornece uma `dynamodb.region` propriedade que você pode definir por tabela do DynamoDB. Quando `dynamodb.region` é definido de forma diferente em duas tabelas, quaisquer dados copiados entre as tabelas ocorrem automaticamente entre as regiões especificadas.

O exemplo a seguir mostra como criar uma tabela do DynamoDB com um script do Hive que define a propriedade `dynamodb.region`:

Note

As propriedades da região por tabela substituem as propriedades globais do Hive.

```
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");
```

Definir valores de throughput do DynamoDB por tabela

O Amazon EMR Hive permite que você defina o `readThroughputPercent` DynamoDB `writeThroughputPercent` e as configurações por tabela na definição da tabela. O script do Amazon EMR Hive a seguir mostra como definir os valores de taxa de transferência. Para obter mais informações sobre valores de throughput do DynamoDB, consulte [Especificar requisitos de leitura e gravação para tabelas](#).

```
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");
```

Configurar um metastore externo para o Hive

Por padrão, o Hive registra as informações do metastore em um SQL banco de dados Meu no sistema de arquivos do nó primário. O metastore contém uma descrição da tabela e dos dados subjacentes nos quais foi criado, incluindo os nomes de partição, os tipos de dados etc. Quando um cluster é encerrado, todos os nós de cluster são desligados, inclusive o nó primário. Quando isso acontece, os dados locais são perdidos porque os sistemas de arquivos dos nós usam

armazenamento temporário. Se precisar que o metastore seja persistido, você deverá criar um metastore externo que exista fora do cluster.

Você tem duas opções para um metastore externo:

- AWS Glue Data Catalog (somente na EMR versão 5.8.0 ou posterior da Amazon).

Para obter mais informações, consulte [Usando o AWS Glue Data Catalog como metastore para o Hive](#).

- Amazon RDS ou Amazon Aurora.

Para obter mais informações, consulte [Usando um SQL banco de dados externo My ou Amazon Aurora](#).

Note

Se você estiver usando o Hive 3 e encontrar muitas conexões com o metastore do Hive, configure o parâmetro `datanucleus.connectionPool.maxPoolSize` para ter um valor menor ou aumente o número de conexões que o servidor de banco de dados pode processar. O aumento do número de conexões se deve à forma como o Hive calcula o número máximo de JDBC conexões. Para calcular o valor ideal de performance, consulte [Propriedades de configuração do Hive](#).

Usando o AWS Glue Data Catalog como metastore para o Hive

Usando a EMR versão 5.8.0 ou posterior da Amazon, você pode configurar o Hive para usar o AWS Glue Data Catalog como seu metastore. Recomendamos essa configuração quando você precisa de um metastore persistente ou de um metastore compartilhado por diferentes clusters, serviços e aplicativos ou contas da AWS .

AWS O Glue é um serviço totalmente gerenciado de extração, transformação e carregamento (ETL) que torna simples e econômico categorizar seus dados, limpá-los, enriquecê-los e movê-los de forma confiável entre vários armazenamentos de dados. O AWS Glue Data Catalog fornece um repositório unificado de metadados em uma variedade de fontes e formatos de dados, com integração com a Amazon, Amazon EMR Redshift, RDS Redshift Spectrum, Athena e qualquer aplicativo compatível com o metastore Apache Hive. AWS Os rastreadores do Glue podem inferir automaticamente o esquema dos dados de origem no Amazon S3 e armazenar os metadados associados no catálogo

de dados. Para obter mais informações sobre o catálogo de dados, consulte [Preenchendo o catálogo de dados do AWS Glue](#) no Guia do desenvolvedor do AWS Glue.

Taxas separadas se aplicam ao AWS Glue. Há uma taxa mensal para armazenar e acessar os metadados no Catálogo de Dados, uma taxa horária cobrada por minuto pelas ETL tarefas do AWS Glue e pelo tempo de execução do rastreador, e uma taxa horária cobrada por minuto para cada endpoint de desenvolvimento provisionado. O Catálogo de Dados permite o armazenamento de até um milhão de objetos gratuitamente. Se você armazenar mais de um milhão de objetos, será cobrado USD \$1 por cada 100.000 objetos acima de um milhão. Um objeto no Catálogo de Dados é uma tabela, uma partição ou um banco de dados. Para obter mais informações, consulte [Definição de preço do Glue](#).

Important

Se você criou tabelas usando o Amazon Athena ou o Amazon Redshift Spectrum antes de 14 de agosto de 2017, bancos de dados e tabelas são armazenados em um catálogo gerenciado pelo Athena, que é separado do Glue Data Catalog. AWS Para integrar a Amazon EMR a essas tabelas, você deve fazer o upgrade para o AWS Glue Data Catalog. Para obter mais informações, consulte [Atualização para o catálogo de dados AWS Glue no Guia](#) do usuário do Amazon Athena.

Especificando o AWS Glue Data Catalog como metastore

Você pode especificar o AWS Glue Data Catalog como metastore usando o AWS Management Console, AWS CLI, ou Amazon. EMR API Ao usar o CLI ou API, você usa a classificação de configuração do Hive para especificar o Catálogo de Dados. Além disso, com o Amazon EMR 5.16.0 e versões posteriores, você pode usar a classificação de configuração para especificar um catálogo de dados em outro. Conta da AWS Ao usar o console, você pode especificar o Catálogo de Dados usando Opções avançadas ou Opções rápidas.

Console

Para especificar o AWS Glue Data Catalog como o metastore do Hive com o console

1. Faça login no AWS Management Console e abra o EMR console da Amazon em <https://console.aws.amazon.com/emr>.
2. EC2 Em EMR Ativado, no painel de navegação esquerdo, escolha Clusters e, em seguida, escolha Criar cluster.

3. Em Pacote de aplicativos, escolha Core Hadoop ou Custom HBase. Se você personalizar seu cluster, certifique-se de selecionar o Hive ou HCatalog como um dos seus aplicativos.
4. Em Configurações do Catálogo de Dados do AWS Glue, marque a caixa de seleção Usar para metadados de tabelas do Hive.
5. Escolha qualquer outra opção que se aplique ao cluster.
6. Para iniciar o cluster, escolha Criar cluster.

CLI

Para especificar o AWS Glue Data Catalog como o metastore do Hive com o AWS CLI

Para obter mais informações sobre como especificar uma classificação de configuração usando o AWS CLI e EMRAPI, consulte [Configurar aplicações](#).

- Especifique o valor para `hive.metastore.client.factory.class` usando a classificação de configuração `hive-site`, conforme mostrado neste exemplo:

```
[
  {
    "Classification": "hive-site",
    "Properties": {
      "hive.metastore.client.factory.class":
      "com.amazonaws.glue.catalog.metastore.AWSGlueDataCatalogHiveClientFactory"
    }
  }
]
```

Nas versões de EMR lançamento 5.28.0, 5.28.1, 5.29.0 ou 6.x, se você estiver criando um cluster usando o AWS Glue Data Catalog como metastore, defina o `hive.metastore.schema.validation false` Isso impede que o Hive and valide o esquema HCatalog do metastore em relação ao My. SQL Sem essa configuração, o grupo de instâncias primárias será suspenso após a reconfiguração no Hive ou. HCatalog

```
[
  {
    "Classification": "hive-site",
    "Properties": {
```

```

    "hive.metastore.client.factory.class":
    "com.amazonaws.glue.catalog.metastore.AWSGlueDataCatalogHiveClientFactory",
    "hive.metastore.schema.validation": "false"
  }
}
]

```

Se você já tem um cluster na EMR versão 5.28.0, 5.28.1 ou 5.29.0, você pode definir o grupo de instâncias primário com as seguintes informações:

`hive.metastore.schema.validation false`

```

Classification = hive-site
Property       = hive.metastore.schema.validation
Value         = false

```

Para especificar um catálogo de dados em uma AWS conta diferente, adicione a `hive.metastore.glue.catalogid` propriedade conforme mostrado no exemplo a seguir. Substitua *acct-id* pela conta da AWS do Catálogo de Dados.

```

[
  {
    "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"
    }
  }
]

```

IAMpermissões

O perfil da EC2 instância de um cluster deve ter IAM permissões para as ações do AWS Glue. Além disso, se você habilitar a criptografia para objetos do AWS Glue Data Catalog, a função

também deverá ter permissão para criptografar, descriptografar e gerar o AWS KMS key usado para criptografia.

Permissões para ações do AWS Glue

Se você usa o perfil de EC2 instância padrão da AmazonEMR, nenhuma ação é necessária. A política `AmazonElasticMapReduceforEC2Role` gerenciada anexada ao `EMR_EC2_DefaultRole` permite todas as ações necessárias do AWS Glue. No entanto, se você especificar um perfil de EC2 instância e permissões personalizados, deverá configurar as ações apropriadas do AWS Glue. Use a política gerenciada `AmazonElasticMapReduceforEC2Role` como ponto de partida. Para obter mais informações, consulte [Função de serviço para EC2 instâncias de cluster \(perfil de EC2 instância\)](#) no Amazon EMR Management Guide.

Permissões para criptografar e descriptografar o Glue Data Catalog AWS

O perfil de instância precisa de permissão para criptografar e descriptografar dados usando a chave. Você não precisa configurar essas permissões se as duas instruções a seguir se aplicarem:

- Você ativa a criptografia para objetos do AWS Glue Data Catalog usando chaves gerenciadas para o AWS Glue.
- Você usa um cluster que está no mesmo catálogo Conta da AWS de dados do AWS Glue.

Caso contrário, você deverá adicionar a seguinte declaração à política de permissões anexada ao seu perfil de EC2 instância.

```
[
  {
    "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"
      }
    ]
  }
]
```


]

Para obter mais informações sobre a criptografia do AWS Glue Data Catalog, consulte [Criptografando seu catálogo de dados](#) no AWS Glue Developer Guide.

Permissões baseadas em recursos

Se você usa o AWS Glue em conjunto com o Hive, o Spark ou o Presto na AmazonEMR, o AWS Glue oferece suporte a políticas baseadas em recursos para controlar o acesso aos recursos do Catálogo de Dados. Esses recursos incluem bancos de dados, tabelas, conexões e funções definidas pelo usuário. Para obter mais informações, consulte [Políticas baseadas em recursos no AWS Glue](#) no Guia do desenvolvedor do AWS Glue.

Ao usar políticas baseadas em recursos para limitar o acesso ao AWS Glue na AmazonEMR, o principal que você especifica na política de permissões deve ser a função ARN associada ao perfil de EC2 instância especificado quando um cluster é criado. Por exemplo, para uma política baseada em recursos anexada a um catálogo, você pode especificar a função da função de serviço padrão ARN para instâncias de clusterEC2, *EMR_EC2_DefaultRole* como oPrincipal, usando o formato mostrado no exemplo a seguir:

```
arn:aws:iam::acct-id:role/EMR_EC2_DefaultRole
```

A ferramenta *acct-id* pode ser diferente do ID da conta AWS Glue. Isso permite o acesso a partir de EMR clusters em contas diferentes. Você pode especificar várias entidades principais, cada uma de uma conta diferente.

Considerações ao usar o Catálogo de Dados do AWS Glue

Considere os seguintes itens ao usar o AWS Glue Data Catalog como metastore com o Hive:

- A adição de auxiliares JARs usando o shell Hive não é suportada. Como solução alternativa, use a classificação de `hive-site` configuração para definir a `hive.aux.jars.path` propriedade, que adiciona auxiliar JARs ao classpath do Hive.
- Não há suporte para [transações do Hive](#).
- Não há suporte para renomear tabelas de dentro do AWS Glue.
- Quando você cria uma tabela do Hive sem especificar um LOCATION, os dados da tabela são armazenados no local especificado pela propriedade `hive.metastore.warehouse.dir`. Por padrão, esse é um local emHDFS. Se outro cluster precisar acessar a tabela, ele apresentará falha, a menos que tenha permissões adequadas para o cluster que criou a tabela. Além disso,

como o HDFS armazenamento é transitório, se o cluster for encerrado, os dados da tabela serão perdidos e a tabela deverá ser recriada. Recomendamos que você especifique a `LOCATION` no Amazon S3 ao criar uma tabela do Hive usando o Glue. AWS Como alternativa, você pode usar a classificação de configuração `hive-site` para especificar um local no Amazon S3 para `hive.metastore.warehouse.dir`, que se aplica a todas as tabelas do Hive. Se uma tabela for criada em um HDFS local e o cluster que a criou ainda estiver em execução, você poderá atualizar a localização da tabela para o Amazon S3 a partir do AWS Glue. Para obter mais informações, consulte Como [trabalhar com tabelas no console do AWS Glue](#) no Guia do desenvolvedor do AWS Glue.

- Valores de partição contendo aspas e apóstrofes não são compatíveis, por exemplo, `PARTITION (owner="Doe ' s")`.
- [As estatísticas de coluna](#) não têm suporte para o emr-5.31.0 e versões posteriores.
- O uso da [autorização do Hive](#) não é compatível. Como alternativa, considere usar as [Políticas baseadas em recursos do AWS Glue](#). Para obter mais informações, consulte [Usar políticas baseadas em recursos para o catálogo de dados Amazon EMR Access to AWS Glue](#).
- Não há suporte para [restrições do Hive](#).
- A [Otimização baseada em custos no Hive](#) não é compatível.
- A configuração de `hive.metastore.partition.inherit.table.properties` não é compatível.
- O uso das seguintes constantes de metastore não é compatível: `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`.
- Quando você usa uma expressão do predicado, os valores explícitos devem estar no lado direito do operador de comparação ou as consultas podem apresentar falha.
 - Correto: `SELECT * FROM mytable WHERE time > 11`
 - Incorreto: `SELECT * FROM mytable WHERE 11 > time`
- EMRAs versões 5.32.0 e 6.3.0 e posteriores da Amazon oferecem suporte ao uso de funções definidas pelo usuário (UDFs) em expressões de predicados. Quando são usadas versões anteriores, suas consultas podem apresentar falha devido à forma como o Hive tenta otimizar a execução da consulta.

- [Tabelas temporárias](#) não são compatíveis.
- Recomendamos criar tabelas usando aplicativos da Amazon EMR em vez de criá-las diretamente usando o AWS Glue. Criar uma tabela por meio do AWS Glue pode fazer com que os campos obrigatórios sejam perdidos e causar exceções de consulta.
- Na EMR versão 5.20.0 ou posterior, a remoção paralela de partições é ativada automaticamente para o Spark e o Hive quando o AWS Glue Data Catalog é usado como metastore. Essa alteração reduz significativamente o tempo de planejamento de consultas ao executar várias solicitações em paralelo para recuperar partições. O número total de segmentos que podem ser executados simultaneamente varia entre 1 e 10. O valor padrão é 5, que é uma configuração recomendada. Você pode alterá-lo especificando a propriedade `aws.glue.partition.num.segments` na classificação de configuração `hive-site`. Se ocorrer controle de utilização, você poderá desativar o atributo alterando o valor para 1. Para obter mais informações, consulte a [Estrutura de segmentos do AWS Glue](#).

Usando um SQL banco de dados externo My ou Amazon Aurora

Para usar um My SQL database externo ou Amazon Aurora como seu metastore do Hive, você substitui os valores de configuração padrão do metastore no Hive para especificar a localização do banco de dados externo, seja em uma instância do Amazon My ou em um Amazon Aurora P. RDS SQL `ostgreSQLinstance`

Note


O Hive não oferece suporte nem impede o acesso simultâneo para gravação em tabelas da metastore. Se você compartilhar informações do metastore entre dois clusters, deverá garantir não gravar na mesma tabela do metastore simultaneamente, a menos que esteja gravando em partições diferentes da mesma tabela do metastore.

O procedimento a seguir mostra como substituir os valores de configuração padrão para a localização da metastore do Hive e iniciar um cluster usando a localização da metastore reconfigurada.

Para criar um metastore localizado fora do cluster EMR

1. Crie um banco de dados My SQL ou Aurora Postgre. SQL Se você usa o PostgreSQL, deve configurá-lo depois de provisionar seu cluster. Somente My SQL é suportado na criação

- do cluster. [Para obter informações sobre as diferenças entre o Aurora My e o SQL Aurora PostgreSQL, consulte Visão geral do Amazon Aurora My e Como trabalhar SQL com o Amazon Aurora Postgre. SQL](#) Para obter informações sobre como criar um RDS banco de dados da Amazon em geral, consulte <https://aws.amazon.com/rds/>.
2. Modifique seus grupos de segurança para permitir JDBC conexões entre seu banco de dados e o grupo de segurança ElasticMapReduce-Master. Para obter informações sobre como modificar seus grupos de segurança para acesso, consulte [Trabalho com grupos de segurança EMR gerenciados pela Amazon](#).
 3. Defina os valores de JDBC configuração em `hive-site.xml`:

 Important

Se você fornecer informações confidenciais, como senhas, para a EMR configuração da AmazonAPI, essas informações serão exibidas para as contas que têm permissões suficientes. Se você teme que essas informações possam ser exibidas para outros usuários, crie o cluster com uma conta administrativa e limite outros usuários (IAMusuários ou aqueles com credenciais delegadas) a acessar os serviços no cluster criando uma função que negue explicitamente as permissões para a chave. `elasticmapreduce:DescribeCluster` API

- a. Crie um arquivo de configuração chamado `hiveConfiguration.json`, que contém edições em `hive-site.xml`, conforme mostrado no exemplo a seguir.

Substituir *hostname* com o DNS endereço da sua RDS instância Amazon executando o banco de dados, e *username* e *password* com as credenciais do seu banco de dados. Para obter mais informações sobre como se conectar às instâncias de banco de dados My SQL e Aurora, consulte [Conectando-se a uma instância de banco de dados executando o mecanismo de SQL banco de dados My](#) e [Conectando-se a um cluster de banco de dados Athena no Guia do usuário](#) da Amazon. `RDS`
`javax.jdo.option.ConnectionURL` é a string de JDBC conexão de uma JDBC metastore. `javax.jdo.option.ConnectionDriverName` é o nome da classe do driver para uma JDBC metastore.

Os Meus SQL JDBC drivers são instalados pela AmazonEMR.

A propriedade "value" não pode conter espaços ou retornos de carro. Eles devem aparecer todos em uma única linha.

```
[
  {
    "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. Faça referência ao `hiveConfiguration.json` arquivo ao criar o cluster, conforme mostrado no AWS CLI comando a seguir. Neste comando, o arquivo é armazenado localmente, você também pode carregar o arquivo para o Amazon S3 e consultá-lo, por exemplo, `s3://DOC-EXAMPLE-BUCKET/hiveConfiguration.json`.

Note

Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

```
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. Conecte-se ao nó primário do cluster.

Para obter informações sobre como se conectar ao nó primário, consulte [Conecte-se ao nó primário usando SSH](#) o Amazon EMR Management Guide.

5. Crie tabelas do Hive especificando o local no Amazon S3 ao digitar um comando semelhante ao seguinte:

```
CREATE EXTERNAL TABLE IF NOT EXISTS table_name
(
  key int,
  value int
)
LOCATION s3://DOC-EXAMPLE-BUCKET/hdfs/
```

6. Adicione seu script do Hive ao cluster em execução.

Seu cluster Hive é executado usando o metastore localizado na Amazon. RDS Execute todos os clusters Hive adicionais que compartilham essa metastore, especificando a localização da metastore.

Usar o driver JDBC do Hive

Você pode usar ferramentas conhecidas de business intelligence, como Microsoft Excel, MicroStrategy, QlikView e Tableau, com o Amazon EMR para explorar e visualizar seus dados. Muitas dessas ferramentas exigem um driver JDBC (Java Database Connectivity) ou ODBC (Open Database Connectivity). O Amazon EMR é compatível com as conectividades do JDBC e do ODBC.

O exemplo abaixo demonstra o uso do SQL Workbench/J como um cliente SQL para conexão a um cluster do Hive no Amazon EMR. Para drivers adicionais, consulte [Usar ferramentas de inteligência comercial com o Amazon EMR](#).

Antes de instalar e trabalhar com o SQL Workbench/J, faça download do pacote de driver e instale o driver. Os drivers incluídos no pacote são compatíveis com as versões do Hive disponíveis nas versões 4.0 e posteriores do Amazon EMR. Para obter notas de release e documentação detalhadas, consulte a documentação do PDF incluída no pacote.

- Download do pacote de drivers JDBC do Hive mais recente

<http://awssupportdatasvcs.com/bootstrap-actions/Simba/latest/>

- Versões mais antigas do driver JDBC do Hive

<http://awssupportdatasvcs.com/bootstrap-actions/Simba/>

Para instalar e configurar o SQL Workbench

1. Faça o download do cliente do SQL Workbench/J para o seu sistema operacional de <http://www.sql-workbench.net/downloads.html>.
2. Instale o SQL Workbench/J. Para obter mais informações, consulte [Instalar e iniciar o SQL Workbench/J](#) (apenas em inglês) no SQL Workbench/J User's Manual.
3. Usuários do Linux, Unix, Mac OS X: em uma sessão de terminal, crie um túnel de SSH para o nó principal do cluster usando o comando a seguir. Substitua *master-public-dns-name* pelo nome DNS público do nó principal e o *path-to-key-file* pelo local e nome do arquivo da chave privada do Amazon EC2 (.pem).

```
ssh -o ServerAliveInterval=10 -i path-to-key-file -N -L 10000:localhost:10000  
hadoop@master-public-dns-name
```

Usuários do Windows: em uma sessão do PuTTY, crie um túnel SSH para o nó principal do cluster (usando o encaminhamento de porta local) com 10000 para Porta de origem e *master-public-dns-name*:10000 para Destino. Substitua *master-public-dns-name* pelo nome DNS público do nó principal.

4. Adicione o driver do JDBC ao SQL Workbench.
 - a. Na caixa de diálogo Select Connection Profile (Selecionar perfil de conexão), clique em Manage Drivers (Gerenciar drivers).
 - b. Clique no ícone Create a new entry (Criar uma nova entrada) (página em branco).
 - c. No campo Name (Nome), digite **Hive JDBC**.
 - d. Em Library (Biblioteca), clique no ícone Select the JAR file(s) (Selecionar os arquivos JAR).
 - e. Navegue até o local que contém os drivers extraídos. Selecione os drivers incluídos na versão do pacote de drivers JDBC que você baixou e clique em Abrir.

Por exemplo, seu pacote de drivers JDBC pode incluir os JARs a seguir.

```
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. Na caixa de diálogo Please select one driver (Selecione um driver), selecione `com.amazon.hive.jdbc41.HS2DriverOK`.
5. Quando você retornar para a caixa de diálogo Manage Drivers (Gerenciar drivers), verifique se o campo Classname (Nome da classe) está preenchido e selecione OK.
6. Quando você retornar para a caixa de diálogo Select Connection Profile (Selecionar perfil de conexão), verifique se o campo Driver está definido como Hive JDBC (JDBC do Hive) e forneça a string de conexão do JDBC no campo URL: `jdbc:hive2://localhost:10000/default`.
7. Selecione OK para conectar. Depois que a conexão estiver concluída, os detalhes da conexão irão aparecer na parte superior da janela do SQL Workbench/J.

Para obter mais informações sobre como usar o Hive e a interface do JDBC, consulte [HiveClient](#) e [HiveJDBCInterface](#) na documentação do Apache Hive.

Melhorar a performance do Hive

A Amazon EMR oferece recursos para ajudar a otimizar o desempenho ao usar o Hive para consultar, ler e gravar dados salvos no Amazon S3.

O S3 Select pode melhorar o desempenho de consultas CSV e JSON arquivos em alguns aplicativos “transferindo” o processamento para o Amazon S3.

O committer otimizado do EMRFS S3 é uma alternativa à [OutputCommitter](#) classe, que elimina as operações de lista e renomeação para melhorar o desempenho ao gravar arquivos que o Amazon S3 usa. EMRFS

Tópicos

- [Habilitando o committer otimizado do Hive EMRFS S3](#)
- [Usar o S3 Select com o Hive para melhorar a performance](#)
- [MSCKOtimização](#)

Habilitando o committer otimizado do Hive EMRFS S3

O Hive EMRFS S3 Optimized Committer é uma forma alternativa de o EMR Hive gravar arquivos para inserir consultas durante o uso. EMRFS O confirmador elimina as operações de listar e renomear feitas no Amazon S3 e melhora a performance da aplicação. O recurso está disponível a partir de EMR 5.34 e EMR 6.5.

Habilitar o confirmador

Se você quiser permitir que o EMR Hive use `HiveEMRFSOptimizedCommitter` para confirmar dados como padrão para todas as tabelas gerenciadas e externas do Hive, use a seguinte `hive-site` configuração nos clusters EMR 6.5.0 ou EMR 5.34.0.

```
[
  {
    "classification": "hive-site",
    "properties": {
      "hive.blobstore.use.output-committer": "true"
    }
  }
]
```

Note

Não ative esse atributo quando `hive.exec.parallel` estiver definido como `true`.

Limitações

As restrições básicas a seguir se aplicam a tags:

- Não há suporte para habilitar o Hive para mesclar arquivos pequenos automaticamente. A lógica de confirmação padrão do Hive será usada mesmo quando o confirmador otimizado estiver habilitado.
- ACIDAs tabelas do Hive não são suportadas. A lógica de confirmação padrão do Hive será usada mesmo quando o confirmador otimizado estiver habilitado.
- A nomenclatura de nomeação para arquivos gravados foi alterada de `<task_id>_<attempt_id>_<copy_n>` para

<task_id>_<attempt_id>_<copy_n>_<query_id> do Hive. Por exemplo, um arquivo denominado

s3://warehouse/table/partition=1/000000_0 será alterado

para s3://warehouse/table/partition=1/000000_0-

hadoop_20210714130459_ba7c23ec-5695-4947-9d98-8a40ef759222-1. query_id Aqui está uma combinação do nome de usuário, carimbo de data/hora e UUID

- Quando partições personalizadas estão em sistemas de arquivos diferentes (HDFS, S3), esse recurso é automaticamente desativado. A lógica de confirmação padrão do Hive será usada quando estiver habilitada.

Usar o S3 Select com o Hive para melhorar a performance

Com a EMR versão 5.18.0 e posterior da Amazon, você pode usar o [S3 Select](#) com o Hive na Amazon. O S3 Select possibilita que as aplicações recuperem apenas um subconjunto dos dados de um objeto. Para a AmazonEMR, o trabalho computacional de filtrar grandes conjuntos de dados para processamento é “transferido” do cluster para o Amazon S3, o que pode melhorar o desempenho em alguns aplicativos e reduzir a quantidade de dados transferidos entre a Amazon e o Amazon S3. EMR

O S3 Select é compatível com tabelas do Hive baseadas em JSON arquivos CSV e definindo a variável de `s3select.filter` configuração como `true` durante sua sessão do Hive. Para ter mais informações e exemplos, consulte [Especificar o S3 Select no código](#).

O S3 Select é adequado para minha aplicação?

Recomendamos que você avalie seus aplicativos com e sem o S3 Seleccione para ver se o uso pode ser adequado para o aplicativo.

Use as seguintes diretrizes para determinar se o seu aplicativo é adequado para o uso do S3 Select:

- Sua consulta filtra mais de metade do conjunto de dados original.
- Os predicados do filtro de consulta usam colunas que têm um tipo de dados compatível com o Amazon S3 Select. Para obter mais informações, consulte [Tipos de dados](#) no Guia do usuário do Amazon Simple Storage Service.
- Sua conexão de rede entre o Amazon S3 e o EMR cluster da Amazon tem boa velocidade de transferência e largura de banda disponível. O Amazon S3 não compacta HTTP respostas, então é provável que o tamanho da resposta aumente para arquivos de entrada compactados.

Considerações e limitações

- A criptografia do lado do servidor do Amazon S3 com chaves de criptografia fornecidas pelo cliente (SSE-C) e a criptografia do lado do cliente não são suportadas.
- A propriedade `AllowQuotedRecordDelimiters` não é compatível. Se essa propriedade for especificada, a consulta falhará.
- Somente JSON arquivos CSV e arquivos no formato UTF -8 são suportados. Várias linhas CSVs e não JSON são suportadas.
- Somente arquivos descompactados, gzip ou bzip2 são compatíveis.
- Caracteres de comentário na última linha não são compatíveis.
- Linhas vazias no final de um arquivo não são processadas.
- O Hive na Amazon EMR oferece suporte aos tipos de dados primitivos compatíveis com o S3 Select. Para obter mais informações, consulte [Tipos de dados](#) no Guia do usuário do Amazon Simple Storage Service.

Especificar o S3 Select no código

Para usar o S3 Select em uma tabela do Hive, crie a tabela especificando com `amazonaws.emr.s3select.hive.S3SelectableTextInputFormat` como o nome da classe `INPUTFORMAT` e especifique um valor para a propriedade `s3select.format` usando a cláusula `TBLPROPERTIES`.

Por padrão, o S3 Select está desabilitado quando você executa consultas. Habilite o S3 Select definindo `s3select.filter` como `true` em sua sessão do Hive, conforme mostrado a seguir. Os exemplos abaixo demonstram como especificar o S3 Select ao criar uma tabela a partir de JSON arquivos subjacentes CSV e, em seguida, consultar a tabela usando uma instrução de seleção simples.

Example CREATETABLEdeclaração para tabela CSV baseada

```
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"
);
```

Example CREATETABLEdeclaração para tabela JSON baseada

```
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"
);
```

Example SELECTTABLEdeclaração

```
SET s3select.filter=true;
SELECT * FROM mys3selecttable WHERE col2 > 10;
```

MSCKOtimização

O Hive armazena uma lista de partições para cada tabela no seu metastore. No entanto, quando as partições são adicionadas ou removidas diretamente do sistema de arquivos, o metastore do Hive não toma conhecimento dessas alterações. O [MSCKcomando](#) atualiza os metadados da partição no metastore do Hive para partições que foram adicionadas ou removidas diretamente do sistema de arquivos. A sintaxe do comando é:

```
MSCK [REPAIR] TABLE table_name [ADD/DROP/SYNC PARTITIONS];
```

O Hive implementa esse comando da seguinte forma:

1. O Hive recupera todas as partições da tabela do metastore. Em seguida, com base na lista de caminhos de partição que não existem no sistema de arquivos, ele cria uma lista de partições a serem descartadas do metastore.
2. O Hive reúne os caminhos de partição presentes no sistema de arquivos, compara-os com a lista de partições do metastore e gera uma lista de partições que precisam ser adicionadas ao metastore.
3. O Hive atualiza a metastore usando o modo ADD, DROP ou SYNC.

Note

Quando há muitas partições no metastore, a etapa para verificar se uma partição não existe no sistema de arquivos leva muito tempo para ser executada porque a `exists` API chamada do sistema de arquivos deve ser feita para cada partição.

Na Amazon EMR 6.5.0, o Hive introduziu uma bandeira chamada.

`hive.emr.optimize.msck.fs.check` Quando ativado, esse sinalizador faz com que o Hive verifique a presença de uma partição na lista de caminhos de partição do sistema de arquivos gerada na etapa 2 acima, em vez de fazer API chamadas ao sistema de arquivos. No Amazon EMR 6.8.0, o Hive habilitou essa otimização por padrão, eliminando a necessidade de definir a bandeira.

`hive.emr.optimize.msck.fs.check`

Usando o Hive Live Long and Process () LLAP

O Amazon EMR 6.0.0 oferece suporte às funcionalidades Live Long and Process (LLAP) para o Hive. LLAP usa daemons persistentes com cache inteligente na memória para melhorar o desempenho da consulta em comparação com o modo de execução de contêiner Tez padrão anterior.

Os LLAP daemons do Hive são gerenciados e executados como um serviço. YARN Como um YARN serviço pode ser considerado um YARN aplicativo de longa execução, alguns dos recursos do cluster são dedicados ao Hive LLAP e não podem ser usados para outras cargas de trabalho. Para obter mais informações, consulte [LLAP](#) e [YARN Serviço API](#).

Habilite o Hive LLAP na Amazon EMR

Para habilitar o Hive LLAP na AmazonEMR, forneça a seguinte configuração ao iniciar um cluster.

```
[
  {
    "Classification": "hive",
    "Properties": {
      "hive.llap.enabled": "true"
    }
  }
]
```

Para obter mais informações, consulte [Configurar aplicações](#).

Por padrão, a Amazon EMR aloca cerca de 60% dos YARN recursos do cluster para daemons do HiveLLAP. Você pode configurar a porcentagem de YARN recursos de cluster alocados para o Hive LLAP e o número de tarefas e nós principais a serem considerados para a alocação do HiveLLAP.

Por exemplo, a configuração a seguir inicia o Hive LLAP com três daemons em três tarefas ou nós principais e aloca 40% dos recursos dos três núcleos ou dos nós de tarefas para os daemons do Hive. YARN LLAP

```
[
  {
    "Classification": "hive",
    "Properties": {
      "hive.llap.enabled": "true",
      "hive.llap.percent-allocation": "0.4",
      "hive.llap.num-instances": "3"
    }
  }
]
```

Você pode usar as seguintes `hive-site` configurações na classificação API para substituir as configurações padrão de LLAP recursos.

Propriedade	Descrição
<code>hive.llap.daemon.yarn.container.mb</code>	Tamanho total LLAP do contêiner do daemon (em MB)

Propriedade	Descrição
hive.llap.daemon.memory.per.instance.mb	A memória total usada pelos executores no contêiner do LLAP daemon (em MB)
hive.llap.io.memory.size	Tamanho do cache para LLAP entrada/saída
hive.llap.daemon.num.executors	Número de executores por daemon LLAP

Inicie o Hive LLAP em seu cluster manualmente

Todas as dependências e configurações usadas pelo LLAP são empacotadas no arquivo LLAP tar como parte da inicialização do cluster. Se LLAP estiver habilitado usando "hive.llap.enabled": "true", recomendamos que você use a EMR reconfiguração da Amazon para fazer LLAP alterações na configuração.

Caso contrário, para qualquer alteração manualhive-site.xml, você deve reconstruir o arquivo LLAP tar usando o `hive --service llap` comando, conforme demonstra o exemplo a seguir.

```
# 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
```

Verifique o status do Hive LLAP

Use o comando a seguir para verificar o status do Hive LLAP por meio do Hive.

```
hive --service llapstatus
```

Use o comando a seguir para verificar o status de LLAP uso YARN do Hive.

```
yarn app -status (name-of-llap-service)
```

```
# example:
```

```
yarn app -status llap0 | jq
```

Iniciar ou parar o Hive LLAP

Como o Hive LLAP é executado como um YARN serviço persistente, você interrompe ou reinicia o YARN serviço para interromper ou reiniciar o HiveLLAP. Os comandos a seguir demonstram isso.

```
yarn app -stop llap0  
yarn app -start llap0
```

Redimensionar o número de daemons do Hive LLAP

Use o comando a seguir para reduzir o número de LLAP instâncias.

```
yarn app -flex llap0 -component llap -1
```

Para obter mais informações, consulte [Flex a component of a service](#).

Criptografia no Hive

Esta seção descreve os tipos de criptografia EMR compatíveis com a Amazon.

Criptografia modular em Parquet no Hive

A criptografia modular em Parquet fornece controle de acesso e criptografia em nível colunar para aprimorar a privacidade e a integridade dos dados armazenados no formato de arquivo Parquet. Esse recurso está disponível no Amazon EMR Hive a partir da versão 6.6.0.

As soluções anteriormente suportadas para segurança e integridade, que incluem a criptografia de arquivos ou a criptografia da camada de armazenamento, estão descritas em [Opções de criptografia](#) no Amazon EMR Management Guide. Essas soluções podem ser usadas para arquivos Parquet, mas o uso dos novos atributos do mecanismo de criptografia integrado do Parquet fornece acesso granular ao nível de coluna, além de melhorias na performance e na segurança. Saiba mais sobre esse atributo na página do Apache no github [Criptografia modular em Parquet](#).

Os usuários transmitem configurações para leitores e gravadores do Parquet usando as configurações do Hadoop. As configurações detalhadas para que os usuários configurem leitores e gravadores para habilitar a criptografia, bem como alternar recursos avançados, estão documentadas em [PARQUET-1854: Interface baseada em propriedades](#) para o gerenciamento de criptografia de parquet

Exemplos de uso

O exemplo a seguir aborda a criação e gravação em uma tabela do Hive usando AWS KMS para gerenciar chaves de criptografia.

1. Implemente um KmsClient para o AWS KMS serviço conforme descrito no documento [PARQUET-1373: Ferramentas de gerenciamento de chaves de criptografia](#). O exemplo a seguir mostra um trecho de implementação.

```
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(masterKeyId).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. Crie suas chaves de AWS KMS criptografia para o rodapé e para as colunas com suas IAM funções tendo acesso conforme descrito em [Criação de chaves](#) no Guia do AWS Key Management Service desenvolvedor. A IAM função padrão é EMR _ ECS _default.
3. No aplicativo Hive em um EMR cluster da Amazon, adicione o cliente acima usando a ADD JAR declaração, conforme descrito na documentação do [Apache Hive](#) Resources. Este é um exemplo de instrução:

```
ADD JAR 's3://location-to-custom-jar';
```

Um método alternativo é adicionar o JAR ao auxlib do Hive usando uma ação de bootstrap. Veja a seguir um exemplo de linha a ser adicionada à ação de bootstrap:

```
aws s3 cp 's3://location-to-custom-jar' /usr/lib/hive/auxlib
```

4. Defina as seguintes configurações:

```

set
parquet.crypto.factory.class=org.apache.parquet.crypto.keytools.PropertiesDrivenCryptoFactory;
set
parquet.encryption.kms.client.class=org.apache.parquet.crypto.keytools.AwsKmsClient;

```

5. Crie uma tabela do Hive com o formato Parquet, especifique as AWS KMS chaves SERDEPROPERTIES e insira alguns dados nela:

```
CREATE TABLE my_table(name STRING, credit_card STRING)
ROW FORMAT SERDE 'org.apache.hadoop.hive ql.io.parquet.serde.ParquetHiveSerDe'
WITH SERDEPROPERTIES (
  'parquet.encrypted.column.key'=<aws-kms-key-id-for-column-1>: credit_card',
  'parquet.encrypted.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. Verifique se ao criar uma tabela externa no mesmo local sem acesso às AWS KMS chaves (por exemplo, acesso à IAM função negado), você não consegue ler os dados.

```
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. A última instrução deve gerar a seguinte exceção:

```
Failed with exception
java.io.IOException:org.apache.parquet.crypto.KeyAccessDeniedException: Footer key:
access denied
```

Criptografia em trânsito em 2 HiveServer

A partir da EMR versão 6.9.0 da Amazon, HiveServer 2 (HS2) é SSL habilitado para TLS/como parte da configuração de [Criptografia em trânsito em 2 HiveServer](#) segurança. Isso afeta a forma como você se conecta à HS2 execução em um EMR cluster da Amazon com a criptografia em trânsito ativada. Para se conectar HS2, você deve modificar os valores dos TRUSTSTORE_PASSWORD

parâmetros TRUSTSTORE_PATH e no JDBCURL. Veja URL a seguir um exemplo de uma JDBC conexão HS2 com os parâmetros necessários:

```
jdbc:hive2://HOST_NAME:10000/
default;ssl=true;sslTrustStore=TRUSTSTORE_PATH;trustStorePassword=TRUSTSTORE_PASSWORD
```

Use as instruções apropriadas para criptografia 2 dentro ou fora do cluster 2 abaixo HiveServer.

On-cluster HS2 access

Se você estiver acessando HiveServer 2 usando o cliente Beeline depois de SSH acessar o nó primário, faça referência `/etc/hadoop/conf/ssl-server.xml` para encontrar os valores dos TRUSTSTORE_PASSWORD parâmetros TRUSTSTORE_PATH e usando a configuração `ssl.server.truststore.location` e `ssl.server.truststore.password`

Os exemplos de comandos a seguir podem ajudar você a recuperar essas configurações:

```
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

Se você estiver acessando HiveServer 2 de um cliente fora do EMR cluster da Amazon, você pode usar uma das seguintes abordagens para obter o TRUSTSTORE_PATH e:

TRUSTSTORE_PASSWORD

- Converta o PEM arquivo que foi criado durante a [configuração de segurança](#) em um JKS arquivo e use o mesmo na JDBC conexãoURL. Por exemplo, com openssl e keytool, use os seguintes comandos:

```
openssl pkcs12 -export -in trustedCertificates.pem -inkey privateKey.pem -out
trustedCertificates.p12 -name "certificate"
keytool -importkeystore -srckeystore trustedCertificates.p12 -srcstoretype pkcs12
-destkeystore trustedCertificates.jks
```

- Como alternativa, faça referência a `/etc/hadoop/conf/ssl-server.xml` para encontrar os valores dos parâmetros TRUSTSTORE_PATH e TRUSTSTORE_PASSWORD usando a configuração `ssl.server.truststore.location` e

`ssl.server.truststore.password`. Baixe o arquivo `truststore` para a máquina cliente e use o caminho na máquina cliente como `TRUSTSTORE_PATH`.

Para obter mais informações sobre como acessar aplicativos de um cliente fora do EMR cluster da Amazon, consulte [Usar o JDBC driver do Hive](#).

Histórico de versões do Hive

A tabela a seguir lista a versão do Hive incluída em cada versão de lançamento da AmazonEMR, junto com os componentes instalados com o aplicativo. Para obter as versões dos componentes em cada versão, consulte a seção Versão de componentes da sua versão em [Versões de lançamento do Amazon EMR 7.x](#), [Versões de lançamento do Amazon EMR 6.x](#) ou [Versões de lançamento do Amazon EMR 5.x](#).

Informações de versão do Hive

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o 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-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-5.36.2	2.3.9	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-dist-cp,

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
		<p>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</p>
emr-7.1.0	3.1.3	<p>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</p>

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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
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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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
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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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
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-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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-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-6.10.1	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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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
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-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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-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-6.8.1	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-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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-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-6.7.0	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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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
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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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
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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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
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-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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-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-6.3.0	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-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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-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-6.2.0	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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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
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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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
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-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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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
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-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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-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
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-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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-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
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-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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-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
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-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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-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
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-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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
emr-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-ftpfs-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-5.28.1	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-ftpfs-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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-ftpfs-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-5.27.1	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-ftpfs-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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-ftpfs-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-5.26.0	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-ftpfs-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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-ftpfs-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-5.24.1	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-ftpfs-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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-ftpfs-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-5.23.1	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-ftpfs-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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-ftpfs-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-5.22.0	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-ftpfs-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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-ftpfs-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-5.21.1	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-ftpfs-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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-ftpfs-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-5.20.1	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-ftpfs-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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-ftpfs-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-5.19.1	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-ftpfs-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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-ftpfs-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-5.18.1	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-ftpfs-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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-ftpfs-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-5.17.2	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-ftpfs-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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-ftpfs-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-5.17.0	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-ftpfs-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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
emr-5.16.1	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-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-5.16.0	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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-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-5.15.0	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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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
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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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
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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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
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-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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-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-5.12.1	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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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
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-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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-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-5.11.2	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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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
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-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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-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-5.10.0	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-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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-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-5.9.0	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-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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-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-5.8.2	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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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
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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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
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-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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-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-5.6.0	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-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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-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-5.5.3	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-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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-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-5.5.1	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-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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-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-5.4.1	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-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-resourcemanager, hadoop-yarn-timeline-server, hive-client, hive-hbase, hcatalog-server, hive-server2, mysql-server, tez-on-yarn
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-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn
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-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn
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-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn
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-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn
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-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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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
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-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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-resourcemanager, hadoop-yarn-timeline-server, hive-client, hcatalog-server, hive-server, mysql-server, tez-on-yarn
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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
emr-4.9.6	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-httpps-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
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-httpps-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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
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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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
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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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
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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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
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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
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-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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
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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-httpps-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
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-httpps-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
emr-4.7.0	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-httfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hcatalog-server, hive-server, mysql-server
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-httfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-httplibs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server
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-httplibs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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
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

Gravadora Amazon EMR Release	Versão do Hive	Componentes instalados com o Hive
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-httfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server
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-httfs-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-httfs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hive-client, hive-metastore-server, hive-server, mysql-server

Notas da versão do Hive por versão

Tópicos

- [Amazon EMR 7.1.0 - Notas de lançamento do Hive](#)
- [Amazon EMR 6.15.0 - Notas de lançamento do Hive](#)
- [Amazon EMR 6.14.0 - Notas de lançamento do Hive](#)
- [Amazon EMR 6.13.0 - Notas de lançamento do Hive](#)
- [Amazon EMR 6.12.0 - Notas de lançamento do Hive](#)
- [Amazon EMR 6.11.0 - Notas de lançamento do Hive](#)
- [Amazon EMR 6.10.0 - Notas de lançamento do Hive](#)
- [Amazon EMR 6.9.0 - Notas de lançamento do Hive](#)
- [Amazon EMR 6.8.0 - Notas de lançamento do Hive](#)
- [Amazon EMR 6.7.0 - Notas de lançamento do Hive](#)
- [Amazon EMR 6.6.0 - Notas de lançamento do Hive](#)

Amazon EMR 7.1.0 - Notas de lançamento do Hive

Amazon EMR 7.1.0 - Mudanças no Hive

Tipo	Descrição
Correção de bugs	HIVE-24381 — A entrada de texto compactad o retorna 0 linhas se o cabeçalho/rodapé de ignorar for incluído.
Correção de bugs	HIVE-24190 —LLAP: ShuffleHandler pode retornar DISK _ ERROR _ de EXCEPTION acordo com -4233. TEZ
Correção de bugs	HIVE-23073 — Sombra Netty.
Correção de bugs	HIVE-23073 — Compartilhe Netty e atualize para netty 4.1.48.Final.

Tipo	Descrição
Correção de bugs	HIVE-23148 — O fluxo do cliente externo do Llap é interrompido devido ao sombreamento netty.
Correção de bugs	HIVE-25180 — Atualiza o Netty.
Correção de bugs	HIVE-24524 — LLAP ShuffleHandler: atualize para o Netty4 e remova a dependência do Netty3 do hive sempre que possível.
Correção de bugs	HIVE-28000 — Hive QL: a cláusula “not in” fornece resultados incorretos quando a coerção de tipo não pode ocorrer.
Correção de bugs	HIVE-27993 — O Netty4 ShuffleHandler deve usar 1 thread principal.
Upgrade	Atualiza o Netty para 4.1.100.Final
Upgrade	Atualiza o Jetty para 9.4.53.v20231009
Upgrade	Atualiza o Zookeeper para 3.9.1

Amazon EMR 7.1.0 - Mudanças no Hive

- O Amazon EMR 7.1 atualiza o Hive para o Netty 4.1.100.Final para resolver as vulnerabilidades de segurança no Netty3. Como depende do netty3, o Hive não hive-druid-handler tem o classpath do Hive hive-druid-handler JAR no Amazon 7.1. EMR Uma próxima EMR versão da Amazon o incluirá no classpath do Hive assim que o manipulador Druid suportar 4.1.100.Final ou versões posteriores do Netty. Entre em contato com o AWS suporte se precisar da hive-druid-handler JAR EMR versão 7.1 ou superior da Amazon.

Amazon EMR 6.15.0 - Notas de lançamento do Hive

Amazon EMR 6.15.0 - Alterações no Hive

Tipo	Descrição
Atributo	Support for TEZ-4397 — Para abertura de divisão assíncrona do Tez, o Hive agora suporta as configurações do Tez descritas em. Abertura de divisão assíncrona do Tez
Correção de bugs	HIVE-25400 — Mova a atualização do offset para. <code>BytesColumnVector setValPre allocated</code>
Correção de bugs	HIVE-25190 — Corrige muitas pequenas alocações em. <code>BytesColumnVector</code>
Upgrade	Atualização do Apache Hadoop para 3.3.6.
Upgrade	HIVE-26684 — Atualize para <code>maven-shade-plugin 3.4.1</code> .
Melhoria	Para reduzir o tempo de inicialização do EMR cluster Amazon, remova 15 segundos de tempo de espera do script de HCatalog inicialização.

Amazon EMR 6.14.0 - Notas de lançamento do Hive

Amazon EMR 6.14.0 - Alterações no Hive

Tipo	Descrição
Melhoria	HIVE-26762 : Remova a poda do operando em <code>HiveFilterSetOpTransposeRule</code>

Tipo	Descrição
Correção de bugs	HIVE-27582 : Não armazene em cache o formato de entrada da HBase tabela em FetchOperator
Correção de bugs	HIVE-26452 : NPE ao converter JOIN para uma JOIN coluna referenciada mais MAPJOIN de uma vez
Correção de bugs	HIVE-26416 : AcidUtils.isRawFormatFile() lança para arquivo InvalidProtocolBufferException non-ORC
Correção de bugs	HIVE-26105 : Mostrar colunas mostra valores extras se a coluna Comentários contiver caracteres chineses específicos
Correção de bugs	HIVE-25864 : A otimização de consultas do Hive cria um plano errado para envio de predicados com função de janela
Correção de bugs	HIVE-25224 : Várias INSERT declarações envolvendo tabelas com resultados diferentes bucketing_versions em erro
Correção de bugs	HIVE-24151 : MultiDelimiterSerDe muda os dados se as strings contiverem caracteres non-ASCII
Correção de bugs	HIVE-23606 : (LLAP) Atraso na limpeza de DirectByteBuffer EncodedReaderImpl
Correção de bugs	HIVE-22165 : A sincronização introduzida por HIVE-14296 em SessionManager.closeSession causa alta latência em um servidor Hive ocupado

Tipo	Descrição
Correção de bugs	HIVE-21304 : Torne o uso da versão de compartimento mais robusto

Amazon EMR 6.13.0 - Notas de lançamento do Hive

Amazon EMR 6.13.0 - Alterações no Hive

Tipo	Descrição
Melhoria	Atualizar os scripts do Python para oferecer suporte ao Python3
Melhoria	HIVE-27097 : Melhore a estratégia de repetição para cliente e servidor MetaStore
Correção de bugs	HIVE-21778 :CBO: “A estrutura não é nula” é avaliada como anulável, sempre causando a perda de filtro na consulta
Correção de bugs	HIVE-21009 : Adicionando a capacidade de o usuário definir o usuário de vinculação
Correção de bugs	HIVE-22661 : Falha na compactação em uma tabela não compartimentada com dados carregados no caminho
Correção de bugs	HIVE-19718 : Adicionar partições em massa também busca a tabela para cada partição
Correção de bugs	HIVE-22173 : A consulta com várias vistas laterais trava durante a compilação
Correção de bugs	HIVE-27088 : Resultados incorretos quando junções internas e externas com filtros de pós-junção são mescladas

Tipo	Descrição
Correção de bugs	HIVE-21935 : Vetorização da colmeia: desempenho degradado com a vetorização UDF
Correção de bugs	HIVE-25299 : A conversão de timestamp para tipos de dados numéricos está incorreta para fusos não horários UTC
Correção de bugs	HIVE-24626 :LLAP: os tópicos do leitor podem ficar inativos se todos os encadeamentos do elevador de E/S estiverem ocupados para enfileirar outros leitores com a fila cheia
Correção de bugs	HIVE-27029 : falha na consulta do hive com erro de fechamento do sistema de arquivos, retrabalho feito para -26352 HIVE
Correção de bugs	HIVE-26352 : A verificação de acesso à fila Tez falha com exceção na compactação GSS
Correção de bugs	HIVE-24590 : O registro de operações ainda vazava anexos log4j
Correção de bugs	HIVE-24552 : Possível vazamento ou acúmulo de HMS conexões em loadDynamicPartitions
Correção de bugs	HIVE-27069 : Resultados incorretos com junção de mapa de compartimento
Correção de bugs	HIVE-27344 : Adicione uma verificação nula em #close RecordReaderImpl
Correção de bugs	HIVE-27439 : Support space em Decimal
Correção de bugs	HIVE-27267 : Resultados incorretos ao fazer a junção do mapa de compartimento em coluna decimal com subconsulta

Tipo	Descrição
Correção de bugs	HIVE-21986 : HiveServer Web UI: definindo o Strict-Transport-Security no cabeçalho de resposta padrão
Correção de bugs	HIVE-22148 : Os tokens de delegação do S3A não são adicionados na configuração do trabalho do Compactor.
Correção de bugs	HIVE-22622 : O Hive permite criar uma estrutura com nomes de atributos duplicados
Correção de bugs	HIVE-22008 : O LIKE operador deve corresponder à entrada de várias linhas
Correção de bugs	HIVE-23144 :LLAP: Deixe QueryTracker a limpeza continuar serviceStop
Correção de bugs	HIVE-22391 : NPE ao verificar o cache de resultados da consulta do Hive
Correção de bugs	HIVE-23305 : NullPointerException em LlapTaskSchedulerService addNode decorrência de condições de corrida
Correção de bugs	HIVE-22178 : Parquet lança depois FilterPredicate CastException SchemaEvolution
Correção de bugs	HIVE-21517 : Corrigir AggregateStatsCache
Correção de bugs	HIVE-21825 : Melhore a mensagem de erro do cliente quando o HA ativo/passivo está ativado
Correção de bugs	HIVE-23389 : FilterMergeRule pode levar a AssertionError
Correção de bugs	HIVE-22767 : Beeline não analisa ponto e vírgula nos comentários corretamente

Tipo	Descrição
Correção de bugs	HIVE-22996 : a BasicStats análise deve verificar proativamente se há uma string nula ou vazia
Correção de bugs	HIVE-22808 : HiveRelFieldTrimmer não suporta HiveTableFunctionScan
Correção de bugs	HIVE-22437 : Cache NPE de metadados no bloqueio de LLAP metadados.
Correção de bugs	HIVE-22606 : AvroSerde registra avro.schema.literal abaixo do nível INFO
Correção de bugs	HIVE-22713 : A propagação constante não deve ser feita para a estrutura Join-Fil (*) -RS
Correção de bugs	HIVE-21624 :LLAP: As métricas da CPU no nível do thread estão quebradas
Correção de bugs	HIVE-22815 : reduza a criação desnecessária de objetos do sistema de arquivos no MROutput
Correção de bugs	HIVE-23060 : Falha na consulta com o erro “A expressão de conjuntos de agrupamento não está na chave BY. GROUP Error encountered near token”
Correção de bugs	HIVE-22236 : Falha ao criar visualização selecionando Visualização contendo subconsulta NOT IN
Correção de bugs	HIVE-19886 : Os registros podem ser direcionados para 2 arquivos se —hiveconf hive.log.file for usado

Tipo	Descrição
Correção de bugs	HIVE-20620 : colisões manifestas ao inserir em tabelas MM classificadas em compartimentos com particionamento dinâmico
Correção de bugs	HIVE-14557 : Ponteiro nulo quando ambos e Mapjoin estão habilitados SkewJoin
Correção de bugs	HIVE-20471 : problemas ao obter o caminho padrão do banco de dados
Correção de bugs	HIVE-20598 : Corrigir erros de digitação nos cálculos HiveAlgorithmsUtil
Correção de bugs	HIVE-14737 : Problema ao acessar /logs em uma interface de usuário da Web do Kerberized Hive Server 2
Correção de bugs	HIVE-20733 : G enericUDFOPEqual NS não pode usar = nas descrições do plano
Correção de bugs	HIVE-20848 : Depois de definir a UpdateInputAccessTimeHook consulta, falha com a tabela não encontrada.
Correção de bugs	HIVE-18929 : O método humanReadableInt em HiveStringUtils .java tem uma condição de corrida.
Correção de bugs	HIVE-20841 LLAP: Torne as portas dinâmicas configuráveis
Correção de bugs	HIVE-20930 : VectorCoalesce no FILTER modo in não entra em vigor
Correção de bugs	HIVE-21007 : Semi-adesão + União pode levar a planos errados

Tipo	Descrição
Correção de bugs	HIVE-21074 : A remoção de consultas de tabelas agrupadas do Hive não funciona para a condição IS NOT NULL
Correção de bugs	HIVE-21223 : CachedStore retorna partição nula quando a partição não existe
Correção de bugs	HIVE-19625 : Exceção potencial NPE e oculta real no Hive# copyFiles
Correção de bugs	HIVE-17020 : A deduplicação agressiva do RS pode remover incorretamente o galho da árvore OP
Correção de bugs	HIVE-20168 : ReduceSinkOperator Registro oculto
Correção de bugs	HIVE-20879 : Usar null em uma expressão de projeção leva a CastException
Correção de bugs	HIVE-20888 : TxnHandler: sort () chamado em listas imutáveis
Correção de bugs	HIVE-19948 : não HiveCli está dividindo o comando por ponto e vírgula corretamente se as aspas estiverem dentro da string
Correção de bugs	HIVE-20621 : GetOperationStatus chamado em resultset.next causando lentidão incremental
Correção de bugs	HIVE-20854 : Padrões sensatos: o intervalo de batimentos cardíacos do Zookeeper do Hive é de 20 minutos, mude para 2
Correção de bugs	HIVE-20330 : HCatLoader não é possível lidar com vários InputJobInfo objetos para um trabalho com várias entradas

Tipo	Descrição
Correção de bugs	HIVE-20787 : o MapJoinBytesTableContainer dummyRow estojo não suporta reutilização
Correção de bugs	HIVE-20331 : A consulta com união total, vista lateral e junção falha com “não é possível encontrar o pai no operador filho”
Correção de bugs	HIVE-19968 : UDF exceção não é descartada
Correção de bugs	HIVE-20410 : a substituição de inserção abortada na tabela transacional causa o erro “Não há histórico suficiente disponível para...”
Correção de bugs	HIVE-20059 : O streaming do Hive deve tentar o prefixo de sombra incondicionalmente, com exceção
Correção de bugs	HIVE-19424 : Em NPE MetadataFormatters
Correção de bugs	HIVE-20355 : Limpe o parâmetro de. HiveConnection setSchema
Correção de bugs	HIVE-20858 : O serializador não foi inicializado corretamente com a configuração em Utilitários. createEmptyBuckets
Correção de bugs	HIVE-20424 : a ferramenta de esquema não deve poluir o histórico da beeline
Correção de bugs	HIVE-20338 : LLAP: Force a identificação de arquivo sintética para sistemas de arquivos que têm impls de protocolo com semântica de mutação HDFS POSIX
Correção de bugs	HIVE-11708 : Operadores lógicos aumentam com ClassCastExceptions NULL

Tipo	Descrição
Correção de bugs	HIVE-21082 : EmHPL/SQL, a declaração declare não suporta variável do tipo caractere
Correção de bugs	HIVE-16690 : Configure o Tez cartesian product edge com base no tamanho do cluster LLAP
Correção de bugs	HIVE-21296 : Eliminando a exceção de lançamento da partição varchar
Correção de bugs	HIVE-14516 : OrcInputFormat SplitGenerator. callInternal
Correção de bugs	HIVE-20981 : streaming/vazamentos AbstractRecordWriter HeapMemoryMonitor
Correção de bugs	HIVE-20043 : HiveServer 2: SessionState tem um bloco de sincronização estático em torno de um AtomicBoolean
Correção de bugs	HIVE-20191 : o aplicativo de PreCommit patch não falha se o patch estiver vazio
Correção de bugs	HIVE-20400 : criar tabela deve sempre usar um caminho totalmente qualificado para evitar possíveis ambigüidades no FS
Correção de bugs	Adicione uma verificação nula skewedInfo antes de acessar colunas distorcidas

Amazon EMR 6.12.0 - Notas de lançamento do Hive

Amazon EMR 6.12.0 - Alterações no Hive

Tipo	Descrição
Melhoria	Suporte adicionado para JDK 11 e JDK 17 Runtime
Melhoria	Adicione suporte para consultar nomes de colunas de palavras-chave reservadas e sensíveis a maiúsculas e minúsculas quando usar o S3 Select. Para usar isso, defina a propriedade da tabela no formato "s3select.column.mapping" = " <i>column1:fieldName1</i> , <i>column2:fieldName2</i> , ..."
Melhoria	HIVE-23133 : As operações numéricas podem ter resultados diferentes nos arcos de hardware
Melhoria	HIVE-27145 : Use StrictMath para funções matemáticas restantes como acompanhamento de -23133 HIVE
Correção de bugs	Corrige a incompatibilidade de caracteres curinga em get_partitions_by_filter e get_num_partitions_by_filter causada pela portabilidade -22900 no Hive 6.4.0 HMS APIs HIVE EMR
Correção de bugs	HIVE-26736 : Falha de autorização para visualizações aninhadas com cláusula WITH
Correção de bugs	HIVE-22416 : registros de operação relacionados ao MR ausentes quando a execução paralela está ativada
Correção de bugs	HIVE-19653 : Pressão incorreta de predicados para groupby com conjuntos de agrupamento

Tipo	Descrição
Correção de bugs	HIVE-22094 : falhas nas consultas com: hive.ql.exec.vector. ClassCastException DecimalColumnVector não pode ser convertido em hive.ql.exec.vector.Decimal64 ColumnVector
Correção de bugs	HIVE-26340 : O PTF operador vetorizado falhará se a consulta tiver a função de janela em maiúsculas
Correção de bugs	HIVE-26184 : COLLECT _ SET com GROUP BY é muito lento quando algumas teclas estão muito distorcidas
Correção de bugs	HIVE-26373 : ClassCastException ao ler carimbos de data/hora da tabela com dados Avro HBase
Correção de bugs	HIVE-26388 : ClassCastException quando não há uma coluna do tipo string na tabela de origem da CTAS consulta Atualização HIVE-26172: Hive - Atualize o Ant para 1.10.11 devido a -2021-36373 e -2021-36374 CVE CVE
Correção de bugs	HIVE-26114 : Corrigir a conexão jdbc hiveserve r2 usando o comando dfs com espaço de prefixo causará exceção
Correção de bugs	HIVE-26396 : A função trunc tem um problema com a interceptação de precisão e o resultado tem muitos 0
Correção de bugs	HIVE-26446 : HiveProtoLoggingHook falha ao preencher o TablesWritten campo para tabelas particionadas.

Tipo	Descrição
Correção de bugs	HIVE-26639 : ConstantVectorExpression e não ExplainTask deve depender do conjunto de caracteres padrão
Correção de bugs	HIVE-22670 : ArrayIndexOutOfBoundsException quando o leitor vetorizado é usado para ler um arquivo de parquet
Correção de bugs	HIVE-23607 : Problema de permissão: a criação de exibição em outra exibição foi bem-sucedida, mas a alteração da exibição falhou
Correção de bugs	HIVE-25498 : Consulta com mais de 31 funções distintas de contagem retorna um resultado errado
Correção de bugs	HIVE-25780 : DistinctExpansion cria mais de 64 conjuntos de agrupamentos II
Correção de bugs	HIVE-23868 : Especificação da função de janela: suporte 0 anterior/seguinte
Correção de bugs	HIVE-24539 : a geração do OrcInputFormat esquema deve respeitar o delimitador de coluna
Correção de bugs	HIVE-23476 LLAP: Pré-aloque arenas também para maiúsculas e minúsculas
Correção de bugs	HIVE-25806 : Possível vazamento em LlapCacheAwareFs - Parquet, IO LLAP
Correção de bugs	HIVE-23498 : Desative HTTP o método Trace em ThriftHttpClientService
Correção de bugs	HIVE-25729 : ThriftUnionObjectInspector deve ser notificado quando totalmente iniciado

Tipo	Descrição
Correção de bugs	HIVE-23846 : Evite serialização e desserialização desnecessárias de bitvetores
Correção de bugs	HIVE-24233 : exceto que a subconsulta lança um ponteiro nulo com o cbo desativado
Correção de bugs	HIVE-24276 : HiveServer 2 loggerconf.jsp Cross-Site Scripting () Vulnerabilidade XSS
Correção de bugs	HIVE-25721 : O resultado da junção externa está errado
Correção de bugs	HIVE-25223 : Selecionar com limite retorna nenhuma linha na tabela não nativa
Correção de bugs	HIVE-25794 : CombineHiveRecordReader: declarações de log em um loop levam à pressão da memória
Correção de bugs	HIVE-23602 : Use o Java Concurrent Package para o conjunto de identificadores de operação
Correção de bugs	HIVE-24045 : Nenhum registro relacionado a quando o banco de dados padrão é criado
Correção de bugs	HIVE-24305 : o esquema decimal avro não preenche adequadamente a escala/precisão se o valor estiver entre aspas
Correção de bugs	HIVE-25844 : Erros de desserialização de exceção podem fazer com que o beeline seja encerrado imediatamente
Correção de bugs	HIVE-25040 : A cascata de descarte do banco de dados não pode remover funções persistentes

Tipo	Descrição
Correção de bugs	HIVE-23501 : VectorDeserializeRow quando tipos complexos são convertidos AOOB em tipos primitivos
Correção de bugs	HIVE-23704 : O Thrift HTTP Server não manipula o identificador de autenticação corretamente
Correção de bugs	HIVE-23529 : CTAS é quebrado para uniontype quando row_deserialize
Correção de bugs	HIVE-24144 : A getIdentifierQuote string in HiveDatabaseMetaData retorna um valor incorreto
Correção de bugs	HIVE-23850 : Permitir PPD quando o assunto não for uma coluna com conjuntos de agrupamento presentes
Correção de bugs	HIVE-24036 : Exceção Kryo ao serializar o plano para chamada getSplits UDF
Correção de bugs	HIVE-25919 : ClassCastException ao inserir o predicado da coluna booleana HBaseStorageHandler
Correção de bugs	HIVE-25261 : RetryingHMSHandler deve envolver o MetaException com uma breve descrição do alvo
Correção de bugs	HIVE-24792 : Potencial vazamento de rosca na operação
Correção de bugs	HIVE-23409 : Se a reabertura do TezSession aplicativo falhar ou o serviço Timeline estiver inativo, o formulário padrão será fechado após uma nova TezSession tentativa SessionPool

Tipo	Descrição
Correção de bugs	HIVE-23615 : Não defere ponteiros nulos na classe de comandos Beeline
Correção de bugs	HIVE-24849 : Crie um tempo limite de soquete de tabela externo quando o local tiver um grande número de arquivos (afeta 3.1.2)
Correção de bugs	HIVE-24193 : A consulta de seleção na tabela de ácidos de colmeia renomeada não produz nenhuma saída
Correção de bugs	HIVE-25209 : SELECT consulta com SUM função produzindo resultado inesperado
Correção de bugs	HIVE-23666 : A checkHashMode eficiência é ignorada quando um operador groupby não tem um conjunto de agrupamento
Correção de bugs	HIVE-23873 : A consulta da JDBCStorageHandler tabela do Hive falha quando está desativado NPE CBO
Correção de bugs	HIVE-24149 : HiveStreamingConnection não fecha a conexão HMS
Correção de bugs	HIVE-25561 : A tarefa eliminada não deve confirmar o arquivo. (afeta as versões 2.x e 3.x)
Correção de bugs	HIVE-25683 : Feche o leitor. AcidUtils isRawFormatArquivo
Correção de bugs	HIVE-24294 : TezSessionPool as sessões podem ser lançadas AssertionError
Correção de bugs	HIVE-24182 : Problema de autorização do Ranger com permanente UDFs

Tipo	Descrição
Correção de bugs	HIVE-22805 : A vetorização com matriz ou mapa condicional não é implementada e gera um erro
Correção de bugs	HIVE-22828 : Decimal64: NVL e as CASE declarações convertem implicitamente decimal64 em 128
Correção de bugs	HIVE-21398 : As colunas que têm estatísticas estimadas não devem ser consideradas como chaves exclusivas
Correção de bugs	HIVE-22490 : Adicionar jarros com caracteres especiais em seu caminho gera um erro
Correção de bugs	HIVE-22700 : As compactações podem vazar memória quando não autorizadas
Correção de bugs	HIVE-22053 : O nome da função não é normalizado ao criar a função
Correção de bugs	HIVE-22595 : As inserções de partições dinâmicas falham na tabela Avro com esquema externo
Correção de bugs	HIVE-21795 : A linha de resumo do rollup pode estar ausente quando um mapjoin está acontecendo em uma tabela particionada
Correção de bugs	HIVE-22987 : ClassCastException em quando é nulo VectorCoalesce DataTypePhysicalVariation
Correção de bugs	HIVE-22814 : ArrayIndexOutOfBoundsException na vetorização getData Type PhysicalVariation

Tipo	Descrição
Correção de bugs	HIVE-22523 : O manipulador de erros LlapRecordReader pode ser bloqueado se a fila estiver cheia
Correção de bugs	HIVE-21796 : ArrayWritableObjectInspector .equals pode levar tempo $O(2^{\text{nesting_depth}})$
Correção de bugs	HIVE-22929 : Desempenho: a análise do identificador entre aspas usa Regex descartável via String.replaceAll()
Correção de bugs	HIVE-21641 : O cliente externo Llap retorna colunas decimais em precisão/escala diferentes em comparação com o beeline
Correção de bugs	HIVE-22207 : Tez: SplitGenerator lança NumberFormatException quando “dfs.block.size” no cluster é “128m”
Correção de bugs	HIVE-22114 : falha na consulta de inserção para tabela particionada de inserção somente quando todos os compartimentos estão vazios
Correção de bugs	HIVE-22165 : Sincronização introduzida por -14296 ativadaHIVE. SessionManager closeSession causa alta latência em um servidor hive ocupado
Correção de bugs	HIVE-22744 : TezTask para o vértice com mais de uma borda deve ter memória de classificação proporcional
Correção de bugs	HIVE-22072 : Alterar a tabela para fazer uma alteração na coluna não atualiza as referências de restrições

Tipo	Descrição
Correção de bugs	HIVE-22075 : Corrige a regressão max-reducers=1 de -14200 HIVE
Correção de bugs	HIVE-22527 : Hive on Tez: O trabalho de mesclar arquivos pequenos será enviado para outra fila (fila padrão)
Correção de bugs	HIVE-22816 : QueryCache: Consultas usando visualizações podem ser armazenadas em cache após a expansão CTE
Correção de bugs	HIVE-22733 : Depois de desativar a propriedade do registro de operação no hive, ainda está HS2 salvando o registro de operação
Correção de bugs	HIVE-22699 : A máscara UDFs deve mascarar o valor numérico 0
Correção de bugs	HIVE-23356 : A agregação de hash está sempre desativada durante o processamento de consultas com expressões de conjuntos de agrupamento.
Correção de bugs	HIVE2-1568 : HiveRelOptUtil isRowFilteringO plano deve pular o projeto
Correção de bugs	HIVE-21760 : A otimização do trabalho compartilhado deve ser ignorada para junções SMB
Correção de bugs	HIVE-22712 : O ReExec driver executa e envia a consulta na fila padrão, independentemente da fila definida pelo usuário
Correção de bugs	HIVE-21397 : BloomFilter para hive, a tabela Managed [ACID] não funciona conforme o esperado

Tipo	Descrição
Correção de bugs	HIVE-23011 : O otimizador de trabalho compartilhado deve verificar os predicados residuais ao comparar as junções
Correção de bugs	HIVE-21412 : PostExecOrcFileDump não funciona com tabelas ACID
Correção de bugs	HIVE-22201 : ConvertJoinMapJoin # checkShuffleSize ForLargeTable lança ArrayIndexOutOfBoundsException se nenhuma mesa grande for selecionada
Correção de bugs	HIVE-21971 : HS2 vaza o carregador de classe devido a `ReflectionUtils:: CONSTRUCTOR _CACHE` com funções temporárias + Genérico UDF
Correção de bugs	HIVE-21368 : Vetorização: Decimal64 -> conversão desnecessária HiveDecimal
Correção de bugs	HIVE-25416 : Vazamento de memória do Hive metastore devido a um bug datanucleus-api-jdo
Correção de bugs	HIVE-22219 : Desativar um gerenciador de nós bloqueia a reinicialização do serviço LLAP
Correção de bugs	HIVE-21793 : CBO recupera as estatísticas da coluna mesmo se hive.stats.fetch.column.stats estiver definido como false
Correção de bugs	HIVE-22163 : CBO: A ativação CBO ativa a estimativa de estatísticas, mesmo quando a estimativa está desativada
Correção de bugs	HIVE-18735 : Crie uma tabela como se perdesse o atributo transacional

Tipo	Descrição
Correção de bugs	HIVE-22433 : Hive JDBC Storage Handler: resultados incorretos obtidos de e da fonte de dados BOOLEAN TIMESTAMP DataType JDBC
Correção de bugs	HIVE-1940 : ObjectStore cleanNotificationEvents OutOfMemory em um grande número de eventos pendentes
Correção de bugs	HIVE-20785 : Nome de chave errado no. JDBC DatabaseMetaData getPrimaryKeys método
Correção de bugs	HIVE-16116 : Beeline lança NPE quando beeline.hiveconfvariables= {} em beeline.properties
Correção de bugs	HIVE-20066 : hive.load.data.owner é comparado ao principal completo
Correção de bugs	HIVE-20489 : Explique o plano de interrupções na consulta
Correção de bugs	HIVE-21033 : Esquecer de fechar a operação corta mais 2 saídas HiveServer
Correção de bugs	HIVE-19888 : Aviso enganoso de "METASTORE_FILTER_HOOK será ignorado" de SessionState
Correção de bugs	HIVE-20303 : INSERT OVERWRITE TABLE PARTITION db.table (...) IF NOT EXISTS lança InvalidTableException
Correção de bugs	HIVE-16144 : CompactionInfo não tem iguais/, mas é usado no Set hashCode

Tipo	Descrição
Correção de bugs	HIVE-20818 : As visualizações criadas com uma WHERE subconsulta considerarão as visualizações referenciadas na subconsulta como entrada direta
Correção de bugs	HIVE-21005 :LLAP: Lendo mais faixas por vazamento ZlibCodecs
Correção de bugs	HIVE-20771 : LazyBinarySerDe falha em estruturas vazias.
Correção de bugs	HIVE-18852 : Mensagem de erro enganosa na validação da tabela de alteração
Correção de bugs	HIVE-21124 :HPL/SQL não suporta a afirmação CREATE TABLE LIKE
Correção de bugs	HIVE-20935 : O upload do tarball do pacote llap falha ao causar falha no início do serviço EC2 LLAP
Correção de bugs	HIVE-20409 : HiveACID: atualizar/excluir/ mesclar não limpa o diretório de teste do hdfs
Correção de bugs	HIVE-20570 : A união ALL com hive.optimize.union.remove=true tem um plano incorreto
Correção de bugs	HIVE-20421 : Entidade de caractere ilegal '\ b' em hive-default.xml.template
Correção de bugs	HIVE-19133 : As métricas de desempenho em fases do WebUI não são HS2 exibidas corretamente
Correção de bugs	HIVE-18977 : Listar partições retorna resultados diferentes com e diretamente JDO SQL

Tipo	Descrição
Correção de bugs	HIVE-20034 : Reverta as alterações no tratamento de MetaStore exceções para fins de compatibilidade com versões anteriores
Correção de bugs	HIVE-20672 : O encadeamento de login LlapTaskSchedulerService deve relatar cada intervalo fixo
Correção de bugs	HIVE-12812 : Habilite <code>mapred.input.dir.recursive</code> por padrão para oferecer suporte à união com a função agregada
Correção de bugs	HIVE-20147 : A ingestão de streaming do Hive se contenta com o registro sincronizado
Correção de bugs	HIVE-19203 : Problema de segurança de rosca em HiveMetaStore
Correção de bugs	HIVE-20091 : Tez: Adicione credenciais de segurança para saída FileSinkOperator
Correção de bugs	HIVE-16906 : O Hive ATSHook deve verificar se há <code>yarn.timeline-service.enabled</code> antes de se conectar ao ATS
Correção de bugs	HIVE-20714 : SHOW <code>tblproperties</code> para uma única propriedade retorna o valor na coluna de nome
Correção de bugs	HIVE-24730 : As classes Shims substituem valores de <code>hive-site.xml</code> e <code>tez-site.xml</code> silenciosamente
Correção de bugs	HIVE-22055 : a contagem selecionada fornece um resultado incorreto após carregar dados do arquivo de texto

Amazon EMR 6.11.0 - Notas de lançamento do Hive

Amazon EMR 6.11.0 - Alterações no Hive

Tipo	Descrição
Melhoria	Adicionado suporte para descarte multithreaded de partições em vários segmentos para melhorar a performance do descarte de partições
Melhoria	Compatível com a leitura de arquivos de consulta codificados do Hive
Melhoria	Habilitado o Tez Shuffle Handler por padrão para trabalhos do Hive no Tez
Bug	Foi adicionada uma opção para habilitar a distribuição determinística de chaves para redutores para corrigir resultados incorretos quando hive.groupby.skewindata está habilitado (relatado em -20220) HIVE
Bug	Corrigida falha no cálculo de estatísticas quando o nome da partição padrão é configurado
Bug	Respeite todos os parâmetros de SSL classificação personalizados passados quando SSL configurado imediatamente para HiveServer 2 em um cluster com criptografia em trânsito ativada
Backport	HIVE-23617 : problemas de API de armazenamento corrigidos FindBug
Backport	HIVE-26408 : Vetorização: corrija a desalocação de colunas de rascunho, não reutilize uma criança como saída ConstantVectorExpression

Tipo	Descrição
Backport	HIVE-23614 : Sempre passe para HiveConfig removeTempOr DuplicateFiles
Backport	HIVE-23354 : Remova a verificação de integridade do tamanho do arquivo de compareTempOr DuplicateFiles
Backport	HIVE-20344 : Corrigido PrivilegeSynchronizer para arremesso. SBA AccessControlException Também introduzida a propriedade hive.privilege.synchronizer para desabilitar o sincronizador de privilégios
Backport	HIVE-15826 : Support a configuração de 'serialization.encoding' para todos SerDes
Backport	HIVE-18284 : Correção NPE ao inserir dados com a cláusula 'distribuir por' com otimização de classificação dynpart
Backport	HIVE-24930 : Operador. setDone() o curto-circuito da operação secundária não é usado no caminho de código vetorizado (se == 1) childSize
Backport	HIVE-24523 : O caminho de leitura vetorizado para LazySimpleSerde não respeita o timestamp for SERDEPROPERTIES
Backport	HIVE-23265 : Conjuntos de linhas duplicados são retornados com os conjuntos Limit e Offset
Backport	HIVE-21492 : não VectorizedParquetRecordReader consigo ler o arquivo de parquet gerado usando a ferramenta Thrift/Custom

Tipo	Descrição
Backport	HIVE-22540 : Vetorização: colunas decimal64 não funcionam com. VectorizedBatchUtil makeLikeColumnVetor ()
Backport	HIVE-22588 : Limpe as linhas restantes para o resto dos conjuntos de agrupamento ao alternar o modo agrupamento vetorial
Backport	HIVE-22551 : BytesColumnVector initBuffer deve limpar o vetor e o comprimento de forma consistente
Backport	HIVE-22448 :CBO: Expanda a contagem múltipla distinta com uma chave de agrupamento por
Backport	HIVE-22248 : Corrigir problemas persistentes de estatísticas
Backport	HIVE-22210 : A vetorização pode reutilizar colunas de saída de computação envolvidas na filtragem
Backport	HIVE-21531 : Vetorização: todos os códigos NULL hash não são computados usando Murmur3
Backport	HIVE-20419 : Vetorização: evite a mutação VectorPartitionDesc após ser usado em uma chave de hashmap
Backport	HIVE-19388 : durante a inicialização ClassCastException VectorMapJoinCommonOperator
Backport	HIVE-21584 : Preparação do Java 11: o carregador de classes do sistema não é URLClassLoader

Tipo	Descrição
Backport	HIVE-25107 : O registro do classpath deve estar no nível (#2271) DEBUG
Backport	HIVE-22097 : java.util incompatível. ArrayList para java 11
Backport	HIVE-23938 :LLAP: JDK11 - alguns argumentos jvm relacionados à rotação do arquivo de log do GC não podem mais ser usados
Backport	HIVE-26226 : Exclua jdk.tools dep do hive-meta store em upgrade-acid
Backport	HIVE-17879 : Atualize o plug-in Datanucleus Maven
Backport	HIVE-27004 : DateTimeFormatterBuilder #appendZoneText não pode analisar 'UTC+' em versões Java superiores a 8
Backport	HIVE-16812 : VectorizedOrcAcidRowBatchReader não filtra eventos de exclusão
Backport	HIVE-17917 : VectorizedOrcAcidRowBatchReader computeOffsetAndOtimização do bucket
Backport	HIVE-19985 ACID: Ignore a decodificação das seções ROW __ID para consultas somente para leitura
Backport	HIVE-20635 : VectorizedOrcAcidRowBatchReader não filtra eventos de exclusão dos arquivos originais
Upgrade	Atualizar o Javadoc para 3.3.1
Upgrade	Atualizar o Javassist para 3.24.1-GA

Tipo	Descrição
Upgrade	Atualização apache-directory-server para 2.0.0-M14

Novas configurações

Nome	Classificação	Descrição
hive.metastore.fs.drop.partition.threads	hive-site	Número de threads principais no grupo de threads de partições descartadas.
hive.metastore.fs.drop.partition.keepalive.time	hive-site	Tempo em segundos em que um thread assíncrono de partições descartadas ociosas (do grupo de threads) aguardará a chegada de uma nova tarefa antes de ser encerrado.
hive.metastore.fs.drop.partition.threadpool.max.queue.size	hive-site	Tamanho máximo da fila a ser usada no grupo de threads para descartar partições do sistema de arquivos.
hive.groupby.enable.deterministic.distribution	hive-site	Habilite a distribuição determinística de chaves para redutores. Isso passará um valor de semente constante ao chamar a função rand usada para particionamento aleatório.
hive.privilege.synchronizer	hive-site	Se os privilégios do autorizador externo devem ser sincroniz

Nome	Classificação	Descrição
		ados periodicamente em 2. HiveServer
hive.cli.query.file.encoding	hive-site	Codificação de arquivo para todos os tipos de arquivos de consulta (arquivo de consulta, arquivo de consulta inicial, arquivo rc etc.) fornecidos nos argumentos da cli.
hive.emr.tez.shuffle.enabled	hive-site	Os trabalhos do Hive no Tez agora usam tez_shuffle por padrão em vez de mapreduce_shuffle como Shuffle Handler padrão.

Configurações descontinuadas

As propriedades de configuração a seguir foram descontinuadas como resultado de [HIVE-23354 e não são mais compatíveis com as versões 6.11.0 e posteriores da Amazon. EMR](#)

Nome	Valor padrão
hive.mapred.reduce.tasks.speculative.execution	false
tez.am.speculation.enabled	false

Amazon EMR 6.10.0 - Notas de lançamento do Hive

Amazon EMR 6.10.0 - Alterações no Hive

Tipo	Descrição
Atributo	Ative controles de acesso AWS Lake Formation baseados para consultas do Apache Hive (gravação) via IAM Passthrough (Hive /Steps). CLI API
Melhoria	Desativar a configuração <code>hive.log.explain.output</code> por padrão para reduzir o tamanho do log
Backport	HIVE-26408 : Vetorização: corrija a desalocação de colunas de rascunho, não reutilize uma criança como saída ConstantVectorExpression
Backport	HIVE-22269 : Corrige a contagem incorreta de redutores em consultas de inserção com partição dinâmica devido à falta de estatísticas causada por -20703. HIVE
Backport	HIVE-22891 : Ignorar a PartitionDesc extração CombineHiveRecord para o modo de não execução LLAP
Backport	HIVE-23804 : Adicione o banco de dados padrão para tabelas específicas de estatísticas de colunas no esquema de metastore do Hive para torná-las compatíveis com versões anteriores
Backport	HIVE-25277 : Exclusão lenta de partições do Hive para armazenamentos de objetos em nuvem com custos caros ListFiles

Tipo	Descrição
Backport	HIVE-19202 : CBO falhou devido à entrada. NullPointerException HiveAggregate isBucketedInput()
Backport	HIVE-19048 : A correção de erros do Beeline Initscript é ignorada
Backport	HIVE-21085 : Registro de visualizações materializadas inicia sessão tez não externa
Backport	HIVE-21675 : CREATE VIEW IF NOT EXISTS retorna um erro em vez de “OK” se a visualização já existir. Esta é uma regressão do Hive 2.
Backport	HIVE-21646 : Tez: Evite escapar do contexto TezTasks de registro de threads
Backport	HIVE-22054 : Evite listagem recursiva para verificar se um diretório está vazio
Backport	HIVE-16587 : NPE ao inserir tipos complexos com valores nulos aninhados
Backport	HIVE-22647 : Habilitar o pool de sessões por padrão
Backport	HIVE-13288 : Mensagem de exceção confusa em. DagUtils localizeResource
Backport	HIVE-23870 : Otimize várias conversões de texto em. WritableHiveCharObjectInspector getPrimitiveJavaObjeto/ HiveCharWritable
Backport	HIVE-21498 : Atualize Thrift para 0.13.0
Backport	HIVE-24378 : Os espaços à esquerda e à direita não são removidos antes da conversão decimal

Tipo	Descrição
Backport	HIVE-21341 : Padrões sensatos: hive.server2.idle.operation.timeout e hive.server2.idle.session.timeout são muito altos
Backport	HIVE-22465 : Adicionar configuração ssl em TezConfigurationFactory
Backport	HIVE-24710 : Otimize a PTF iteração para contagem (*) para reduzir o custo de E/S CPU
Backport	HIVE-15406 : Considere vetorizar a nova função 'trunc'
Backport	HIVE-21541 : Corrige os cabeçalhos asf ausentes de -15406 HIVE
Backport	HIVE-24808 : datas analisadas em cache
Backport	HIVE-24746 :PTF: TimestampValueBoundaryScanner pode ser otimizado durante o cálculo do intervalo
Backport	HIVE-25059 : O evento Alter é convertido em renomear durante a replicação
Backport	HIVE-25142 : Refazer o hashing na tabela de hash rápida do mapa causando corrupção em chaves grandes
Backport	HIVE-23756 : Adicionou mais restrições ao arquivo package.jdo
Backport	HIVE-25150 : Os caracteres de tabulação não são removidos antes da conversão decimal semelhante ao caractere de espaço que é fixado como parte de -24378 HIVE

Tipo	Descrição
Backport	HIVE-25093 : <code>date_format ()</code> UDF está retornando a saída somente no fuso horário UTC
Backport	HIVE-25268 : <code>date_format</code> udf retorna resultados incorretos para datas anteriores a 1900 se o fuso horário local for diferente de UTC
Backport	HIVE-25338 : AIOBE na configuração se a entrada estiver vazia UDF
Backport	HIVE-22400 : UDF <code>minuto</code> com retorno de tempo NULL
Backport	HIVE-25058 :PTF: <code>TimestampValueBoundaryScanner</code> pode ser otimizado durante o cálculo do intervalo <code>pt2 - isDistanceGreater</code>
Backport	HIVE-25449 : <code>datediff ()</code> fornece uma saída errada quando executado em uma tarefa tez com algum fuso horário diferente UTC
Backport	HIVE-23688 : Vetorização: <code>IndexArrayOutOfBoundsException</code> para coluna de tipo de mapa que inclui valor nulo
Backport	HIVE-22247 : <code>HiveHFile OutputFormat</code> lança <code>FileNotFoundException</code> quando a saída da tarefa da partição está vazia
Backport	HIVE-25570 : O Hive deve enviar o URL caminho completo para autorização do comando <code>insert overwrite location</code>
Backport	HIVE-22903 : <code>row_number ()</code> vetorizado redefine o número da linha após um lote no caso de expressão constante na cláusula de partição

Tipo	Descrição
Backport	HIVE-25549 : Resultados incorretos para a função de janela com expressão na cláusula BY ou PARTITION BY ORDER
Backport	HIVE-25579 : LOAD sobrescrever anexos em vez de sobrescrever
Backport	HIVE-25659 : As consultas SQL diretas do Metastore com IN/ (NOTIN) devem ser divididas com base nos parâmetros máximos permitidos pelo banco de dados SQL
Backport	HIVE-20502 : Corrija NPE ao executar skewjoin_mapjoin10.q quando as estatísticas da coluna são usadas.
Backport	HIVE-25765 : a propriedade skip.header.line.count ignora as linhas de cada bloco quando o tamanho do arquivo é maior FetchOperator
Bug	Corrige NPE na inserção em determinados cenários quando <code>hive.stats.column.autogather</code> e quando ambos <code>hive.groupby.skewindata</code> estão ativados
Bug	Corrige NPE quando <code>mapred.tasktracker.expiry.interval</code> o valor não está definido

Amazon EMR 6.9.0 - Notas de lançamento do Hive

Amazon EMR 6.9.0 - Alterações no Hive

Tipo	Descrição
Upgrade	Atualizar o Jetty para 9.4.48.v20220622

Tipo	Descrição
Upgrade	Suporte para Hadoop 3.3.3
Atributo	Integração do Amazon EMR Hive com o Lake Formation para uso de cargas de trabalho interativas. GCSC API
Atributo	Integração do Amazon EMR Hive com o Iceberg.
Melhoria	Ative SSL em HiveServer 2 quando a criptografia em trânsito estiver habilitada usando as configurações de EMR segurança da Amazon.
Melhoria	Habilite o committer otimizado Hive EMRFS Amazon S3 por padrão. Para obter mais informações, consulte, Habilitando o committer otimizado do Hive EMRFS S3 .
Melhoria	Adicione HiveHBaseTableInputFormatV2 que herda apenas a versão mapeada de InputFormat para corrigir -34210. SPARK Defina <code>hive.hbase.inputformat.v2</code> como <code>true</code> para usá-lo.
Melhoria	Espere que o TeZam seja iniciado em segundo plano com hive.cli.tez.session.async em vez de encerrá-lo e iniciar um novo imediatamente. Use <code>hive.emr.cli.tez.session.open.timeout</code> para definir esse tempo limite em segundos.
Melhoria	Adicione a opção hive.conf.restricted.list.append para anexar configurações separadas por vírgula à lista de configurações restritas existente. <code>hive.conf.restricted.list</code>

Tipo	Descrição
Melhoria	Mensagem de erro mais clara quando a consulta do Hive apresenta falha porque o local não está definido para o banco de dados.
Backport	HIVE-24484 : Atualize o Hadoop para 3.3.1 e o Tez para 0.10.2
Backport	HIVE-22398 : Remova o gerenciamento de YARN filas via. ShimLoader
Backport	HIVE-23190 :LLAP: modifique IndexCache para passar o objeto do sistema de arquivos para o. TezSpillRecord
Backport	HIVE-22185 : HADOOP -15832 causará problemas com testes usando clusters. MiniYarn
Backport	HIVE-21670 : Substituindo por mockito-all dependência. mockito-core
Backport	HIVE-24542 : Prepare a goiaba para atualizações.
Backport	HIVE-23751 :QTest: Substitua o #mkdirs() método ProxyFileSystem para alinhar após -16582. HADOOP
Backport	HIVE-21603 : Preparação do Java 11: atualize a versão do powermock.
Backport	HIVE-24083 : hcatalog erro no Hadoop 3.3.0: tipo de autenticação necessário.
Backport	HIVE-24282 : Mostrar colunas não deve classificar as colunas de saída, a menos que seja explicitamente mencionado.

Tipo	Descrição
Backport	HIVE-20656 : Padrões sensatos: as configurações de memória de agregação de mapas são muito agressivas.
Backport	HIVE-25443 : A seta SerDe não pode serializar/desserializar tipos de dados complexos quando há mais de 1024 valores
Backport	HIVE-19792 : Atualize o orc para 1.5.2 e habilite os testes de evolução do esquema decimal_64.
Backport	HIVE-20437 : manipula a evolução do esquema a partir de flutuação, dupla e decimal.
Backport	HIVE-21987 : O Hive não consegue ler o Parquet int32 anotado com decimal.
Backport	HIVE-20038 : Atualize consultas em lançament os de tabelas sem compartimentos e particion adas. NPE

Amazon EMR 6.9.0 - Problemas conhecidos do Hive

- Com o Amazon EMR 6.6.0 a 6.9.x, INSERT as consultas com partição dinâmica e uma cláusula ORDER BY ou SORT BY sempre terão dois redutores. Esse problema é causado pela OSS alteração [HIVE-20703](#), que coloca a otimização dinâmica da partição de classificação sob decisão baseada em custos. Se sua workload não exigir a classificação de partições dinâmicas, recomendamos que você defina a propriedade `hive.optimize.sort.dynamic.partition.threshold` como -1 para desabilitar o novo atributo e obter o número de redutores calculado corretamente. Esse problema foi corrigido no OSS Hive como parte do [HIVE-22269](#) e foi corrigido no Amazon 6.10.0. EMR

Amazon EMR 6.8.0 - Notas de lançamento do Hive

Amazon EMR 6.8.0 - Alterações no Hive

Tipo	Descrição
Melhoria	Reduza as chamadas do sistema de arquivos no comando msck. Melhorias de performance (~ 15 a 20 x em mais de 10 mil partições)
Backport	HIVE-20678 : HiveHBase TableOutputFormat deve implementar HiveOutputFormat para garantir a compatibilidade
Backport	HIVE-21040 : msck faz uma listagem desnecessária de arquivos no último nível da árvore de diretórios
Backport	HIVE-21460 : Carregar dados seguidos por uma consulta select * resulta em resultados incorretos
Backport	HIVE-21660 : Resultado errado quando a união completa e a visualização posterior com explosão são usadas
Backport	HIVE-22505 : ClassCastException causado pela seleção incorreta do operador vetorizado
Backport	HIVE-22513 : A propagação constante da coluna moldada em operações de filtro pode causar resultados incorretos
Backport	HIVE-23435 : O resultado completo da junção externa é ausência de linhas
Backport	HIVE-24209 : Conversão incorreta de argumentos de pesquisa para NOT BETWEEN operação quando a vetorização está ativada

Tipo	Descrição
Backport	HIVE-24934 : a VectorizedExpressions anotação não é necessária no G Check enericUDFSQCount
Backport	HIVE-25278 : HiveProjectJoinTransposeRul e pode fazer transformações inválidas com expressões de janela
Backport	HIVE-25505 : Resultados incorretos com cabeçalho. skip.header.line.count se a primeira linha estiver em branco
Backport	HIVE-26080 : Atualize o accumulo-core para 1.10.1
Backport	HIVE-26235 : A condição OR na coluna binária está retornando um resultado vazio
Bug	Corrige vários registros SLF4J de aviso de ligações no stderr durante o lançamento
Bug	Corrige a falha na SHOW TABLE EXTENDED consulta com o erro Wrong FS quando a partição e a tabela estão em sistemas de arquivos diferentes.

Amazon EMR 6.8.0 - Problemas conhecidos do Hive

- Com o Amazon EMR 6.6.0 a 6.9.x, INSERT as consultas com partição dinâmica e uma cláusula ORDER BY ou SORT BY sempre terão dois redutores. Esse problema é causado pela OSS alteração [HIVE-20703](#), que coloca a otimização dinâmica da partição de classificação sob decisão baseada em custos. Se sua workload não exigir a classificação de partições dinâmicas, recomendamos que você defina a propriedade `hive.optimize.sort.dynamic.partition.threshold` como `-1` para desabilitar o novo atributo e obter o número de redutores calculado corretamente. Esse problema foi corrigido no OSS Hive como parte do [HIVE-22269](#) e foi corrigido no Amazon 6.10.0. EMR

Amazon EMR 6.7.0 - Notas de lançamento do Hive

Amazon EMR 6.7.0 - Alterações no Hive

Tipo	Descrição
Atributo	Integração do Amazon EMR Hive com LakeFormation.
Atributo	Registro de auditoria adicional para o comissário otimizado EMRFS do Hive Amazon S3. Configuração do Hive: <code>hive.blobstore.output.committer.logging</code> , padrão: <code>false</code>
Atributo	Excluído o diretório de destino na instrução <code>insert overdrive</code> com um resultado de seleção vazio em uma tabela/partição estática não particionada para se comportar de forma semelhante ao Hive 2.x. Configuração do Hive: <code>hive.emr.iow.clean.target.dir</code> , padrão: <code>false</code>
Bug	Falha de consulta intermitente corrigida ao usar o comissário otimizado Hive EMRFS Amazon S3 com classificação de compartimentos de partições.
Upgrade	Versão do Hive atualizada para 3.1.3. Consulte as notas de lançamento do Apache Hive 3.1.3 para obter mais detalhes.
Upgrade	Atualizado o Parquet para 1.12.2 .
Backport	HIVE-20065 : O Metastore não deve confiar no jackson 1.x
Backport	HIVE-20071 : Migre para jackson 2.x e evite o uso

Tipo	Descrição
Backport	HIVE-20607 : TxnHandler deve ser usado PreparedStatement para executar consultas diretas SQL
Backport	HIVE-20740 : Remova o bloqueio global. ObjectStore setConf método
Backport	HIVE-20961 : Aposentar a implementação NVL
Backport	HIVE-22059 : o jar hive-exec não contém a biblioteca jackson (fasterxml)
Backport	HIVE-22351 : Corrige o uso incorreto do encadeamento em ObjectStore TestObjectStore
Backport	HIVE-23534 : NPE em RetryingMetaStoreClient #invoke ao capturar sem mensagem MetaException
Backport	HIVE-24048 : Harmonize os componentes de Jackson com a versão 2.10.latest - Hive
Backport	HIVE-24768 : Use jackson-bom em todos os lugares para substituir a versão
Backport	HIVE-24816 : Atualize jackson para 2.10.5.1 ou 2.11.0+ devido a -2020-25649 CVE
Backport	HIVE-25971 : O desligamento da tarefa Tez está atrasado devido ao pool de threads em cache não estar fechado
Backport	HIVE-26036 : NPE causado por getMTable () em ObjectStore

Amazon EMR 6.7.0 - Problemas conhecidos do Hive

- Consultas com funções de janela na mesma coluna da junção podem levar a transformações inválidas, conforme relatado em [HIVE-25278](#), e causar resultados incorretos ou falhas na consulta. Uma solução alternativa seria desativar essas consultas CBO no nível da consulta. A correção estará disponível em uma EMR versão da Amazon após a 6.7.0. Para obter mais informações, entre em contato com AWS o suporte.
- Com o Amazon EMR 6.6.0 a 6.9.x, INSERT as consultas com partição dinâmica e uma cláusula ORDER BY ou SORT BY sempre terão dois redutores. Esse problema é causado pela OSS alteração [HIVE-20703](#), que coloca a otimização dinâmica da partição de classificação sob decisão baseada em custos. Se sua workload não exigir a classificação de partições dinâmicas, recomendamos que você defina a propriedade `hive.optimize.sort.dynamic.partition.threshold` como `-1` para desabilitar o novo atributo e obter o número de redutores calculado corretamente. Esse problema foi corrigido no OSS Hive como parte do [HIVE-22269](#) e foi corrigido no Amazon 6.10.0. EMR

Amazon EMR 6.6.0 - Notas de lançamento do Hive

Amazon EMR 6.6.0 - Alterações no Hive

Tipo	Descrição
Upgrade	Atualize o Parquet para 1.12.1 .
Upgrade	Atualizar a versão de jetty jars para 9.4.43.v20210629
Bug	Corrigido um problema que fazia com que o Hive fosse instalado em todos os nós de tarefa/núcleo quando LLAP estava ativado em um cluster do Hive.
Backport	HIVE-25942 : Atualize o commons-io para 2.8.0 devido a -2021-29425 CVE
Backport	HIVE-25726 : Velocidade de atualização para 2,3 devido a -2020-13936 CVE

Tipo	Descrição
Backport	HIVE-25680 : Autorize o #get_table_meta HiveMetastore Server API a usar qualquer modelo de autorização. HiveMetastore
Backport	HIVE-25554 : Atualize a versão da seta para 0,15
Backport	HIVE-25242 : A consulta é executada de forma extremamente lenta com vectorized.adaptor = escolhido
Backport	HIVE-25085 : MetaStore Os clientes não são mais compartilhados entre as sessões.
Backport	HIVE-24827 : A consulta de agregação do Hive retorna resultados incorretos para arquivos que não sejam de texto.
Backport	HIVE-24683 : getFileId Hadoop23Shims propenso a caminhos inexistentes NPE
Backport	HIVE-24656 : CBO falha em consultas com é nulo nos tipos de mapa e matriz
Backport	HIVE-24556 : Otimize DefaultGraphWalker para casos sem neto
Backport	HIVE-24408 : Atualize o Parquet para 1.11.1
Backport	HIVE-24391 : Corrigir FIX TestOrcFile falhas no branch-3.1
Backport	HIVE-24362 : o processamento de AST árvores não é ideal para árvores com grande número de nós

Tipo	Descrição
Backport	HIVE-24316 : Atualização de 1.5.6 para 1.5.8 na ORC ramificação-3.1
Backport	HIVE-24307 : Beeline com arquivo de propriedades e parâmetro -e está falhando
Backport	HIVE-24245 : Vetorizado PTF com contagem e partição distinta, produzindo resultados incorretos.
Backport	HIVE-24224 : Corrige ignorar o cabeçalho/rodapé do Hive on Tez em arquivo compactado
Backport	HIVE-24157 : Modo estrito para falha no CAST timestamp ↔ numérico
Backport	HIVE-24113 : NPE em GenericUDFToUnixTimeStamp
Backport	HIVE-23987 : Atualize a versão da seta para 0.11.0
Backport	HIVE-23972 : Adicionar ID de cliente externo ao cliente externo LLAP
Backport	HIVE-23806 : Evite limpar os estados estatísticos da coluna em todas as partições caso o esquema seja estendido. Isso melhora o runtime da instrução alter table add columns.
Backport	HIVE-23779 : As BasicStatsTask informações não estão sendo impressas no console beeline
Backport	HIVE-23306 : o RESET comando não funciona se houver uma configuração definida pelo sistema. getProperty

Tipo	Descrição
Backport	HIVE-23164 : O servidor não foi encerrado corretamente devido a threads que não são do daemon
Backport	HIVE-22967 : Support hive.reloadable.aux.jars.path para Hive on Tez
Backport	HIVE-22934 : contadores de log interativos do servidor Hive para fluxo de erros
Backport	HIVE-22901 : A substituição de variáveis pode levar a referências não circulares OOM
Backport	HIVE-22769 : resultados de consulta incorretos e falha na consulta durante a geração dividida para arquivos de texto compactados
Backport	HIVE-22716 : A leitura para ByteBuffer é interrompida ParquetFooterInputFromCache
Backport	HIVE-22648 : Atualize o Parquet para 1.11.0
Backport	HIVE-22640 : Decimal64ColumnVector: ClassCastException quando o tipo de coluna de partição é Decimal
Backport	HIVE-22621 : caso de teste instável: TestLlapSignerImpl testSigning
Backport	HIVE-22533 : Corrige possíveis vulnerabilidades da interface do usuário da web do LLAP daemon
Backport	HIVE-22532 : PTFPPD pode ultrapassar o limite incorretamente por meio da classificação/funcão DenseRank

Tipo	Descrição
Backport	HIVE-22514 : HiveProtoLoggingHook pode consumir muita memória
Backport	HIVE-22476 : A função datediff do Hive forneceu resultados inconsistentes quando hive.fetch.task.conversion é definida como none
Backport	HIVE-22429 : Tabelas em cluster migradas usando bucketing_version 1 no hive 3 usam bucketing_version 2 para inserções
Backport	HIVE-22412 : StatsUtils lance ao explicar NPE
Backport	HIVE-22360 : MultiDelimitSerDe retorna resultados errados na última coluna quando o arquivo carregado tem mais colunas do que aquelas no esquema da tabela
Backport	HIVE-22332 : O Hive deve garantir configurações válidas de evolução do esquema desde -540 ORC
Backport	HIVE-22331 : unix_timestamp sem argumento retorna timestamp em milissegundos em vez de segundos
Backport	HIVE-2275 : OperationManager queryIdOperation não limpa corretamente vários queryIds
Backport	HIVE-22273 : Falha na verificação de acesso quando um diretório temporário é removido
Backport	HIVE-22270 : Atualize o commons-io para 2.6

Tipo	Descrição
Backport	HIVE-22241 : Implemente UDF para interpretar data/data e hora usando sua representação interna e o calendário híbrido gregoriano-juliano
Backport	HIVE-22241 : Implemente UDF para interpretar data/data e hora usando sua representação interna e o híbrido gregoriano-juliano
Backport	HIVE-22232 : NPE quando hive.order.columnalignment é definido como false
Backport	HIVE-22231 : A consulta Hive com tamanho grande via Knox falha com tubo quebrado Falha na gravação
Backport	HIVE-22221 : Cliente externo Llap - É necessário reduzir # LlapBaselInputFormat getSplits
Backport	HIVE-22208 : O nome da coluna com a palavra-chave reservada não escapa quando a consulta, incluindo junção na tabela com a coluna de máscara, é reescrita
Backport	HIVE-22197 : Common Merge join lança exceção de conversão de classe.
Backport	HIVE-22170 : from_unixtime e unix_timestamp devem usar o fuso horário da sessão do usuário
Backport	HIVE-22169 : Tez: SplitGenerator tenta procurar arquivos de planos que não existirão para Tez
Backport	HIVE-22168 : Remova registros muito caros do hotpath do cache llap

Tipo	Descrição
Backport	HIVE-2161: FunctionRegistry sincroniza em UDF org.apache.hadoop.hive.ql.udf.UDFTypeclasse
Backport	HIVE-22120 : Corrige ArrayOutOfBounds resultados/exceções errados nas junções do mapa externo esquerdo em condições de limite específicas
Backport	HIVE-22115 : Impeça a criação do anexador de roteamento de consultas se a propriedade for definida como false
Backport	HIVE-22113 : Impedir o LLAP desligamento de itens relacionados AMReporter RuntimeException
Backport	HIVE-22106 : Remova a sincronização de consultas cruzadas para a avaliação de partição
Backport	HIVE-22099 : Várias datas relacionadas não UDFs conseguem lidar com datas julianas adequadamente desde -20007 HIVE
Backport	HIVE-22037 : HS2 deve registrar ao desligar devido a OOM
Backport	HIVE-21976 : O deslocamento deve ser nulo em vez de zero em calcita HiveSortLimit
Backport	HIVE-21924 : Divida arquivos de texto mesmo que exista cabeçalho/rodapé
Backport	HIVE-21913 : G enericUDTFGet Splits deve lidar com nomes de usuário da mesma forma que LLAP

Tipo	Descrição
Backport	HIVE-21905 : Melhorias dos genéricos em toda a classe FetchOperator
Backport	HIVE-21902 : HiveServer 2 UI: o cabeçalho de resposta jetty precisa de X-Frame-Options
Backport	HIVE-21888 : Defina o padrão hive.parquet.timestamp.skip.conversion como verdadeiro
Backport	HIVE-21868 : Vetorizar... CAST FORMAT
Backport	HIVE-21864 : # LlapBaseInputFormat closeAll
Backport	HIVE-21863 : Melhore a conversão do tipo vetorizador para expressão WHEN
Backport	HIVE-21862 : ORC ppd produz resultado errado com carimbo de data/hora
Backport	HIVE-21846 : Crie um tópico no TezAm que busque métricas periodicamente LlapDaemon
Backport	HIVE-21837 : MapJoin está lançando uma exceção quando a coluna selecionada tem valores completamente nulos
Backport	HIVE-21834 : Evite chamadas desnecessárias para simplificar as condições do filtro
Backport	HIVE-21832 : Novas métricas para obter o tempo médio de fila/serviço/resposta
Backport	HIVE-21827 : Várias chamadas em SemanticAnalyzer não passam pelo método getObjectByName
Backport	HIVE-21822 : Exponha LlapDaemon métricas por meio de um novo método API

Tipo	Descrição
Backport	HIVE-21818 :CBO: A cópia TableRelOptHiveTable tem tráfego de metastore
Backport	HIVE-21815 : As estatísticas no ORC arquivo são analisadas duas vezes
Backport	HIVE-21805 : HiveServer 2: Use o rápido ShutdownHookManager APIs
Backport	HIVE-21799 : NullPointerException em DynamicPartitionPruningOptimization, quando a chave de junção está na coluna de agregação
Backport	HIVE-21794 : Adicionar parâmetros de visualização materializada a sqlStdAuth SafeVarNameRegexes
Backport	HIVE-21768 :JDBC: Remove o prefixo de união padrão para consultas não incluídas UNION
Backport	HIVE-21746 : ArrayIndexOutOfBoundsException durante a junção de hash particionada dinamicamente, com desativado CBO
Backport	HIVE-21717 : A renomeação do diretório está falhando na tarefa de movimentação.
Backport	HIVE-21685 : Simplificação errada na consulta com várias cláusulas IN
Backport	HIVE-21681 : Descrever formatado mostra informações incorretas para várias chaves primárias
Backport	HIVE-21651 : Mova as sementes protobuf para o hive-exec.

Tipo	Descrição
Backport	HIVE-21619 : Imprima o tipo de carimbo de data/hora sem precisão na explicação estendida SQL
Backport	HIVE-21592 : não OptimizedSql é mostrado quando a expressão contém CONCAT
Backport	HIVE-21576 : Introduzir... CAST FORMATE lista limitada de:formatos de data e hora de 2016 SQL
Backport	HIVE-21573 : O transporte binário deve ignorar o principal se a autenticação estiver definida como delegationToken
Backport	HIVE-21550 : TestObjectStore os testes estão instáveis - Não foi possível obter um bloqueio dentro do tempo solicitado
Backport	HIVE-21544 : A propagação constante corrompe as expressões coalesce/case/when durante o dobramento
Backport	HIVE-21539 : GroupBy + onde a cláusula na mesma coluna resulta em reescrita incorreta da consulta
Backport	HIVE-21538 : Beeline: fonte da senha, embora o leitor do console não tenha passado para o parâmetro de conexão
Backport	HIVE-21509 : LLAP pode armazenar em cache vetores de coluna corrompidos e retornar um resultado de consulta errado

Tipo	Descrição
Backport	HIVE-21499 : não deve remover a função do registro se o comando de criação falhar com AlreadyExistsException
Backport	HIVE-21496 : O dimensionamento automático de buffer não ordenado pode transbordar
Backport	HIVE-21468 : Diferenciação de maiúsculas e minúsculas em nomes de identificadores para manipulador de armazenamento JDBC
Backport	HIVE-21467 : Remova as importações obsoletas de <code>JUNIT.framework.Assert</code>
Backport	HIVE-21435 : <code>LlapBaselInputFormat</code> deve obter o número da tarefa de <code>TASK _ ATTEMPT _ID</code> conf, se presente, durante a construção <code>SubmitWorkRequestProto</code>
Backport	HIVE-21389 : A distribuição Hive perdeu <code>javax.ws.rs-api.jar</code> após -21247 HIVE
Backport	HIVE-21385 : Permitir a desativação do envio de computação não divisível para fontes JDBC
Backport	HIVE-21383 : manipulador JDBC de armazenamento: use catálogo e esquema para recuperar tabelas, se especificado
Backport	HIVE-21382 : Otimização de redução de grupos por chaves - as chaves não são reduzidas na consulta ²³
Backport	HIVE-21362 : Adicione um formato de entrada e sirva para ler os arquivos protobuf.

Tipo	Descrição
Backport	HIVE-21340 :CBO: Remove colunas não chave que alimentam um SemiJoin
Backport	HIVE-21332 : Limpe os buffers não bloqueados em vez dos bloqueados
Backport	HIVE-21329 : Tamanho do buffer de saída não ordenado em tempo de execução do Tez personalizado, dependendo do pipeline do operador
Backport	HIVE-21295 : StorageHandler deve converter a data em string usando a convenção Hive
Backport	HIVE-21294 : Vetorização: o 1-reducer Shuffle pode ignorar as funções de hash do objeto
Backport	HIVE-21255 : Remover em QueryConditionBuilder JdbcStorageHandler
Backport	HIVE-21253 : Support em DB2 JDBC StorageHandler
Backport	HIVE-21232 :LLAP: Adicionar um provedor de afinidade dividido compatível com cache e sem perda
Backport	HIVE-21214 : MoveTask : Use attemptId em vez do tamanho do arquivo para deduplicação de arquivos compareTempOr DuplicateFiles
Backport	HIVE-21184 : Adicione, explique e explique o plano formatado com informações CBO de custo
Backport	HIVE-21182 : Ignore a configuração do diretório de rascunho do hive durante o planejamento

Tipo	Descrição
Backport	HIVE-21171 : Ignore a criação de pastas de rascunho para chá, se estiver ativado RPC
Backport	HIVE-21126 : Permitir consultas em nível de sessão em # LlapBaselInputFormat getSplit
Backport	HIVE-21107 : Erro “Não é possível encontrar o campo” durante a junção de hash particionada dinamicamente
Backport	HIVE-21061 : falha na CTAS consulta com IllegalStateException uma fonte vazia
Backport	HIVE-21041 : NPE, ParseException na obtenção do esquema do plano lógico
Backport	HIVE-21013 : JdbcStorageHandler falha ao encontrar a coluna de partição no Oracle
Backport	HIVE-21006 : Estenda SharedWorkOptimizer para remover semijunções quando houver uma oportunidade de reutilização
Backport	HIVE-20992 : Divida a configuração hive.meta store.dbaccess.ssl.properties em configurações mais significativas
Backport	HIVE-20989 : JDBC - O log GetOperationStatus + pode bloquear o progresso da consulta por meio do modo de suspensão
Backport	HIVE-20988 : resultados incorretos para consultas de grupo por com chave primária em várias colunas

Tipo	Descrição
Backport	HIVE-20985 : Se as entradas do operador selecionado forem colunas temporárias, a vetorização poderá reutilizar algumas delas como saída
Backport	HIVE-20978 : “hive.jdbc.*” deve ser adicionado a sqlStdAuth SafeVarNameRegexes
Backport	HIVE-20953 : Remova uma função do registro de funções quando ela não puder ser adicionada ao metastore ao criá-la.
Backport	HIVE-20952 : Limpando .java VectorizationContext
Backport	HIVE-20951 :LLAP: Defina Xms para 50% sempre
Backport	HIVE-20949 : Melhorar a estimativa de PKFK cardinalidade no planejamento físico
Backport	HIVE-20944 : Não valida as estatísticas durante a compilação da consulta
Backport	HIVE-20940 : Casos intermediários em que a resolução de tipo do Calcite é mais rígida do que o Hive.
Backport	HIVE-20937 : Falha na consulta jdbc do Postgres com “não deve ser negativo” LIMIT
Backport	HIVE-20926 : A dica de redução de semi-junção falha quando as entradas do filtro Bloom estão altas ou quando não há estatísticas
Backport	HIVE-20920 : Use SQL restrições para melhorar o algoritmo de reordenação de junções

Tipo	Descrição
Backport	HIVE-20918 : Sinalizador para ativar/desativar o envio da computação do Calcite para a conexão JDBC
Backport	HIVE-20915 : Disponibilize a otimização dinâmica da partição de classificação para HoS e MR
Backport	HIVE-20910 : A inserção na tabela agrupada falha devido à otimização dinâmica da classificação de partições
Backport	HIVE-20899 : O Keytab URI for LLAP YARN Service é restritivo apenas ao suporte HDFS
Backport	HIVE-20898 : Para funções relacionadas ao tempo, os argumentos não podem ser convertidos em um tipo não anulável
Backport	HIVE-20881 : A propagação constante simplifica demais as projeções
Backport	HIVE-20880 : Atualize o valor padrão para <code>hive.stats.filter.in.min.ratio</code>
Backport	HIVE-20873 : Use o hash Murmur para reduzir a colisão de hash <code>VectorHashKeyWrapperTwoLong</code>
Backport	HIVE-20868 : A SMB junção falha intermitentemente quando o filho está entrando <code>TezDummyOperator</code> <code>getFinalOp</code> <code>MapRecordProcessor</code>
Backport	HIVE-20853 : Expor. <code>ShuffleHandler</code> <code>registerDaemonIlapAPI</code>

Tipo	Descrição
Backport	HIVE-20850 : Envie a condição de maiúsculas e minúsculas das projeções para as tabelas de dimensões, se possível
Backport	HIVE-20842 : corrige a lógica introduzida em HIVE -20660 para estimar estatísticas para o grupo por
Backport	HIVE-20839 : Erro “Não é possível encontrar o campo” durante a junção de hash particionada dinamicamente
Backport	HIVE-20835 : A interação entre restrições e reescrita de MV pode criar um loop no planejador Calcite
Backport	HIVE-20834 : QueryResultCache entradas do Hive mantendo referência a uma consulta em cache SemanticAnalyzer
Backport	HIVE-20830 : falha na declaração JdbcStorageHandler de consulta de intervalo em alguns casos
Backport	HIVE-20829 : JdbcStorageHandler lances divididos de alcance NPE
Backport	HIVE-20827 : Resultados inconsistentes para matrizes vazias
Backport	HIVE-20826 : HiveSemiJoin Regra aprimorada para converter junção + grupo do lado esquerdo para semijunção esquerda
Backport	HIVE-20821 : Reescreva SUM 0 na combinação + SUM COALESCE

Tipo	Descrição
Backport	HIVE-20815 : JdbcRecordReader .next não deve comer exceção
Backport	HIVE-20813 : udf to_epoch_milli também precisa suportar timestamp sem fuso horário.
Backport	HIVE-20804 : Outras melhorias no agrupamento por otimização com restrições
Backport	HIVE-20792 : Inserir carimbo de data/hora com zonas trunca os dados
Backport	HIVE-20788 : A redução estendida de SJ pode retroceder colunas incorretamente ao criar filtros
Backport	HIVE-20778 : A reordenação de junções pode não ser acionada se todas as junções no plano forem criadas pela lógica de decorrelação
Backport	HIVE-20772 : registra contadores por tarefa em CPU LLAP
Backport	HIVE-20768 : Adicionando janela giratória UDF
Backport	HIVE-20767 : Vários projetos entre operadores de junção podem afetar a reordenação da junção usando restrições
Backport	HIVE-20762 : NOTIFICATION _ o intervalo de LOG limpeza está codificado como 60s e é muito pequeno
Backport	HIVE-20761 : Selecione para atualização na tabela notification_sequence com intervalo de repetição e contagem de tentativas muito pequena

Tipo	Descrição
Backport	HIVE-20751 : Atualize a versão da seta para 0.10.0
Backport	HIVE-20746 : HiveProtoHookLogger não fecha o arquivo no final do dia.
Backport	HIVE-20744 : Use SQL restrições para melhorar o algoritmo de reordenação de junções
Backport	HIVE-20740 : Remova o bloqueio global. ObjectStore setConf método. Este cherrypick backports HIVE -20740 é destinado ao Hive 3.2 e 4.x a 3.1.x
Backport	HIVE-20734 : Beeline: Quando o beeline-s ite.xml está e o hive CLI redireciona para o beeline, ele deve usar o nome de usuário/s enha fictícia do sistema em vez de solicitar um
Backport	HIVE-20731 : o arquivo de armazenamento de chaves deve ser autorizado JdbcStorageHandler
Backport	HIVE-20720 : Adicionar opção de coluna de partição ao manipulador JDBC
Backport	HIVE-20719 : a SELECT instrução falha depois UPDATE com a otimização e vetorização de hive.optimize.sort.dynamic.partition ativadas
Backport	HIVE-20718 : Adicionar driver perf cli com restrições
Backport	HIVE-20716 : defina o valor padrão para hive.cbo.stats.correlated.multi.key.joins como true

Tipo	Descrição
Backport	HIVE-20712 : HivePointLookupOptimizer deve extrair casos profundos
Backport	HIVE-20710 : A dobra constante não pode criar constantes nulas sem tipos
Backport	HIVE-20706 : external_jdbc_table2.q falhando intermitentemente
Backport	HIVE-20704 : Estenda HivePreFilteringRule para oferecer suporte a outras funções
Backport	HIVE-20703 : Coloque a otimização dinâmica da partição de classificação sob decisão baseada em custos
Backport	HIVE-20702 : contabilize a sobrecarga das estimativas com reconhecimento da estrutura de dados durante a seleção do mapjoin
Backport	HIVE-20692 : Habilitar o dobramento de expressões NOT x IS (NOT) [TRUE] FALSE
Backport	HIVE-20691 : Corrija org.apache.hadoop.hive.cli. TestMiniLlapCliDriver. testCliDriver[cttl]
Backport	HIVE-20682 : A execução de consulta assíncrona pode falhar se o compartilhamento sessionHive for fechado pelo thread mestre
Backport	HIVE-20676 : HiveServer 2: não Privilege Synchronizer está definido para o status daemon
Backport	HIVE-20660 : A estimativa de agrupamento por estatísticas pode ser melhorada limitando o número total de linhas à tabela de origem

Tipo	Descrição
Backport	HIVE-20652 : junção JdbcStorageHandler push de duas fontes de dados diferentes ao driver jdbc
Backport	HIVE-20651 : a JdbcStorageHandler senha deve ser criptografada
Backport	HIVE-20649 : gerenciador de memória LLAP consciente para escritores Orc
Backport	HIVE-20648 :LLAP: O grupo de vetores por operador deve usar memória por executor
Backport	HIVE-20646 : A condição do filtro de partição não é transferida para a consulta do metastore se tiver IS NOT NULL
Backport	HIVE-20644 : Evite expor informações confidenciais por meio de uma exceção do Hive Runtime
Backport	HIVE-20636 : Melhore a estimativa do número de valores nulos após a junção externa
Backport	HIVE-20632 : A consulta com get_splits UDF falha se a visualização materializada for criada na tabela consultada
Backport	HIVE-20627 : consultas assíncronas simultâneas falham intermitentemente e causam vazamento de memória LockException
Backport	HIVE-20623 : Trabalho compartilhado: estenda o compartilhamento de entradas de cache de junção de mapas em LLAP

Tipo	Descrição
Backport	HIVE-20619 : Incluir MultiDelimitSerDe em HiveServer 2 por padrão
Backport	HIVE-20618 : Durante a seleção de junções, BucketMapJoin pode ser escolhido para tabelas sem compartimentos
Backport	HIVE-20617 : Corrija o tipo de constantes nas expressões IN para ter o tipo correto
Backport	HIVE-20612 : Crie um novo sinalizador de correlação de junção de várias chaves para CBO
Backport	HIVE-20603 : Erro “FS errado” ao inserir na partição após alterar o sistema de arquivos de localização da tabela
Backport	HIVE-20601 : EnvironmentContext nulo em _evento em ALTER PARTITION DbNotificationListener
Backport	HIVE-20583 : Use o nome de host canônico somente para autenticação kerberos em HiveConnection
Backport	HIVE-20582 : Torne o registro de protótipos hflush no hive configurável
Backport	HIVE-20563 : Vetorização: a CASE WHEN expressão falha quando o THEN ELSE tipo/e o tipo de resultado são diferentes
Backport	HIVE-20558 : Altere o padrão de hive.hash table.key.count.adjustment para 0,99

Tipo	Descrição
Backport	HIVE-20552 : Obtenha o esquema de uma forma mais rápida LogicalPlan
Backport	HIVE-20550 : Alterne WebHCat para usar o beeline para enviar consultas do Hive
Backport	HIVE-20537 : Várias colunas unem estimativas com colunas não correlacionadas diferentes no Hive CBO
Backport	HIVE-20524 : A verificação da evolução do esquema foi interrompida ao passar da versão 2 do Hive para a versão 3 para ALTER TABLE VARCHAR DECIMAL
Backport	HIVE-20522 : HiveFilterSetOpTransposeRule pode gerar erro de afirmação devido à nulidade dos campos
Backport	HIVE-20521 : HS2 doAs =true tem problema de permissão com hadoop.tmp.dir, com sistema de arquivos MR e S3A
Backport	HIVE-20515 : resultados de consulta vazios ao usar cache de resultados e diretório temporário de consulta, diretório de cache de resultados em sistemas de arquivos diferentes
Backport	HIVE-20508 : O Hive não suporta nomes de usuário do tipo “user @realm”
Backport	HIVE-20507 : Beeline: adicione um comando utilitário para recuperar todos os uris do beeline-site.xml
Backport	HIVE-20505 : atualize org.openjdk.jmh:jmh-core para 1.21

Tipo	Descrição
Backport	HIVE-20503 : Use estimativas com reconhecimento de estrutura de dados durante a seleção de junção de mapas
Backport	HIVE-20498 : Tipo de data de suporte para coleta automática de estatísticas de colunas
Backport	HIVE-20496 : Vetorização: vetorizada PTF IllegalStateException
Backport	HIVE-20494 : GenericUDFRestrict Information onSchema está quebrado após -19440 HIVE
Backport	HIVE-20477 : não OptimizedSql é mostrado se a expressão contiver INs
Backport	HIVE-20467 : Permitir IF /IF na NOT EXISTS criação/eliminação do plano de recursos EXISTS
Backport	HIVE-20462 : "CREATEVIEWIF NOTEXISTS" falhará se a visualização já existir
Backport	HIVE-20455 : Velocidade de log de security. authorization. PrivilegeSynchronizer.correr
Backport	HIVE-20439 : Use o limite de memória inflado durante a seleção da junção para llap
Backport	HIVE-20433 : A conversão implícita de string em timestamp é lenta
Backport	HIVE-20432 : Reescreva BETWEEN para IN para tipos inteiros para estimativa de estatísticas

Tipo	Descrição
Backport	HIVE-20423 : Definir NULLS LAST como a ordem nula padrão
Backport	HIVE-20418 : LLAP O IO não pode manipular ORC arquivos com índice de linha desativado o corretamente para consultas sem colunas selecionadas
Backport	HIVE-20412 : em NPE HiveMetaHook
Backport	HIVE-20406 : Nested Coalesce fornecendo resultados incorretos
Backport	HIVE-20399 : com CTAS um local de tabela personalizado que não seja totalmente qualificado, falha nas tabelas MM
Backport	HIVE-20393 : Redução de semijunção: For se comporta de forma inconsistente markSemiJoin DPP
Backport	HIVE-20391 : HiveAggregateReduceFunction sRule pode inferir tipo de retorno errado ao decompor a função agregada
Backport	HIVE-20383 : Nome de fila inválido e problemas de sincronização no gancho de eventos proto do hive.
Backport	HIVE-20367 : Vetorização: Support streaming para,,, PTF AVG MAX MIN SUM
Backport	HIVE-20366 : As estimativas de status da consulta TPC -DS 78 estão desativadas para seu filtro nulo

Tipo	Descrição
Backport	HIVE-20364 : Atualize o padrão para hive.map.aggr.hash.min.reduction
Backport	HIVE-20352 : Vetorização: função de agrupamento Support
Backport	HIVE-20347 : hive.optimize.sort.dynamic.partition deve funcionar com particionado e MV CTAS
Backport	HIVE-20345 : O banco de dados Drop pode travar se as tabelas forem excluídas de uma chamada diferente
Backport	HIVE-20343 : Hive 3 : CTAS não respeita as propriedades_transacionais
Backport	HIVE-20340 : Druid Needs Explicit CASTs desde o timestamp até STRING quando a saída da função timestamp é usada como Strin
Backport	HIVE-20339 : Vetorização: levante restrições desnecessárias, fazendo com que algumas não sejam vetorizadas PTF RANK
Backport	HIVE-20337 : CachedStore: getPartitionsBy Expr não está preenchendo a lista de partições corretamente
Backport	HIVE-20336 : Políticas de mascaramento e filtragem para visualizações materializadas
Backport	HIVE-20326 : Crie restrições com RELY como padrão em vez de NO RELY

Tipo	Descrição
Backport	HIVE-20321 : Vetorização: reduza o tamanho da memória de 1 coluna para <1 VectorHas hKeyWrapper CacheLine
Backport	HIVE-20320 : ative a sinalização hive.optimize.remove.sq_count_check
Backport	HIVE-20315 : Vetorização: corrija mais problemas com resultados errados e evite NULL conversões/conversões desnecessárias
Backport	HIVE-20314 : Incluir a remoção de partições na reescrita da visualização materializada
Backport	HIVE-20312 : Permitir que os clientes do Arrow usem seus próprios com BufferAllocator LlapOutputFormatService
Backport	HIVE-20302 LLAP: execução não vetorizada em IO ignora colunas virtuais, incluindo __ID ROW
Backport	HIVE-2030 : VectorFileSinkArrowOperator
Backport	HIVE-20299 : corrida potencial no teste unitário de signatários LLAP
Backport	HIVE-20296 : Melhore HivePointLookupOptimizerRule para poder extrair de contextos mais sofisticados
Backport	HIVE-20294 : Vetorização: problemas de correção/resultados incorretos emNULL/COALESCE/ELT
Backport	HIVE-20292 : Ordem de junção incorreta na consulta tpcds 93 com restrição primária definida

Tipo	Descrição
Backport	HIVE-20290 : Inicialização lenta para que não ArrowColumnarBatchSerDe aloque buffers durante GetSplits
Backport	HIVE-20281 : SharedWorkOptimizer falha com 'o conteúdo do cache do operador e a diferença real do plano'
Backport	HIVE-20277 : Vetorização: expressões de caso que retornam não são suportadas para BOOLEAN FILTER
Backport	HIVE-20267 : Expandindo o WebUI para incluir um formulário para configurar dinamicamente os níveis de log
Backport	HIVE-20263 : Erro de digitação na variável HiveReduceExpressionsWithStatsRule
Backport	HIVE-20260 : NDV de uma coluna não deve ser dimensionada quando a contagem de linhas é alterada pelo filtro em outra coluna
Backport	HIVE-20252 : Redução de semijunção: os ciclos devido à ramificação de semijunção podem permanecer indetectados se um lado pequeno da tabela tiver uma junção de mapa a montante.
Backport	HIVE-20245 : Vetorização: problemas de correção/resultados incorretos em NULL /IN BETWEEN
Backport	HIVE-20241 : Especificação de particionamento de suporte em declarações CTAS

Tipo	Descrição
Backport	HIVE-20240 : Redução de semijunção: use a variável local para verificar a condição da tabela externa
Backport	HIVE-20226 : HMS getNextNotification lançará uma exceção quando a solicitação maxEvents exceder as max_rows da tabela
Backport	HIVE-20225 : SerDe para suportar o formato binário Teradata
Backport	HIVE-20213 : Atualize o Calcite para 1.17.0
Backport	HIVE-20212 : Hiveserver2 no modo http emitindo a métrica default.general.op_en_connections incorretamente
Backport	HIVE-20210 : O otimizador Simple Fetch deve fazer MapReduce com que o filtro na coluna sem partição e a conversão sejam mínimos
Backport	HIVE-20209 : A conexão do Metastore falha na primeira tentativa no repl dump
Backport	HIVE-20207 : Vetorização: Corrigir/Problemas de resultados incorretos no NULL filtro/comparação
Backport	HIVE-20204 : conversão de tipo durante IN
Backport	HIVE-20203 : A seta vaza um SerDe DirectByteBuffer
Backport	HIVE-20197 : Vetorização: adicione testes DECIMAL_64, adicione aritmética de data/intervalo/timestamp e adicione mais agregação BY GROUP

Tipo	Descrição
Backport	HIVE-20193 : não cboInfo está presente no plano de explicação json
Backport	HIVE-20192 : HS2 com metastore incorporado está vazando objetos JDOPersistenceManager
Backport	HIVE-20183 : A inserção de uma tabela com compartimentos pode causar perda de dados, se a tabela de origem contiver um compartimento vazio
Backport	HIVE-20177 : Vetorização: reduza KeyWrapper a alocação no modo Streaming GroupBy
Backport	HIVE-20174 : Vetorização: problemas de NULL correção/resultados incorretos nas funções de agregação BY GROUP
Backport	HIVE-20172 : StatsUpdater falha com GSS exceção ao tentar se conectar ao metastore remoto
Backport	HIVE-20153 : Count and Sum UDF consomem mais memória no Hive 2+
Backport	HIVE-20152 : redefina o estado do banco de dados, quando o repl dump falhar, para que a renomeação da tabela possa ser feita
Backport	HIVE-20149 : TestHiveCli falha/tempo limite
Backport	HIVE-20130 : Melhor registro para sincronizador de esquema de informações
Backport	HIVE-20129 : reverter para a evolução do esquema baseada em posição para tabelas orc

Tipo	Descrição
Backport	HIVE-2018 : SessionStateUserAuthenticator getGroupNames
Backport	HIVE-20116 : TezTask está usando o registrador principal
Backport	HIVE-20115 : Tabelas ácidas não devem usar digitalização de rodapé para análise
Backport	HIVE-20103 : WM: somente DAG contadores agregados se pelo menos um for usado
Backport	HIVE-20101 : BloomKFilter: Evite usar inteiramente as matrizes locais de bytes []
Backport	HIVE-20100 : OpTraits : Selecione Optraits que deve parar quando uma incompatibilidade for detectada
Backport	HIVE-20098 : Estatísticas: NPE ao obter estatísticas de partição da coluna Data
Backport	HIVE-20095 : Corrige o recurso para enviar a computação para tabelas externas do jdbc
Backport	HIVE-20093 LlapOutputFormatService: Use ArrowBuf com Netty para contabilidade
Backport	HIVE-20090 : Estenda a criação de filtros de redução de semijunção para poder descobrir novas oportunidades
Backport	HIVE-20088 : O caminho de localização da configuração do Beeline está montado incorretamente

Tipo	Descrição
Backport	HIVE-20082 : HiveDecimal a conversão para string não formata o decimal corretamente
Backport	HIVE-20069 : Corrige a reotimização em caso de otimização de DPP semijunção
Backport	HIVE-20051 : Ignorar autorização para tabelas temporárias
Backport	HIVE-20044 : O Arrow Serde deve preencher valores de caracteres e manipular cadeias de caracteres vazias corretamente
Backport	HIVE-20028 : A configuração do cache do cliente Metastore é usada incorretamente
Backport	HIVE-20025 : Limpeza de arquivos de eventos criados por HiveProtoLoggingHook
Backport	HIVE-20020 : O frasco de contribuição do Hive não deve estar na biblioteca
Backport	HIVE-20013 : Adicione uma conversão implícita ao tipo de data para a função to_date
Backport	HIVE-20011 : Saia do modo de anexação no gancho de registro de proto
Backport	HIVE-20005 : acid_table_stats, acid_no_buckets, etc - alteração do resultado da consulta na ramificação
Backport	HIVE-20004 : Escala errada usada por ConvertDecimal 64 ToDecimal resulta em resultados incorretos

Tipo	Descrição
Backport	HIVE-19995 : tráfego de linhas agregado para tabelas de ácidos
Backport	HIVE-19993 : Não é possível usar um alias de tabela que também aparece como nome de coluna
Backport	HIVE-19992 : Vetorização: continuação de -19951 --> adicionar chamada a. HIVE SchemaEvolution isOnlyImplicitConversão para desativar a LLAP E/S codificada ORC somente quando a conversão do tipo de dados não estiver implícita
Backport	HIVE-19989 : O Metastore usa um nome de aplicativo errado para métricas HADOOP2
Backport	HIVE-19981 : Tabelas gerenciadas convertidas em tabelas externas pelo HiveStrictManagedMigration utilitário devem ser configuradas para excluir dados quando a tabela for descartada
Backport	HIVE-19967 : SMB Junte-se: Need Optraits for ala Op PTFOperator GBY
Backport	HIVE-19935 : Sessão do Hive WM encerrada: falha ao atualizar a contagem de tarefas LLAP
Backport	HIVE-19924 : Tag: tarefas distcp executadas pelo Repl Load
Backport	HIVE-19891 : inserir em tabelas externas com diretórios de partições personalizados pode causar perda de dados

Tipo	Descrição
Backport	HIVE-19850 : A remoção dinâmica de partições no Tez está levando ao erro “Nenhum trabalho encontrado para o tablesan”
Backport	HIVE-19806 : Classifique a saída de qtests para evitar descamação nos resultados do teste
Backport	HIVE-19770 : Support CBO para consultas com várias colunas iguais em select
Backport	HIVE-19769 : Crie objetos dedicados para nomes de banco de dados e tabelas
Backport	HIVE-19765 : Adicione testes específicos do Parquet ao BlobstoreCliDriver
Backport	HIVE-19759 : Teste escamoso: # TestRpc testServerPort
Backport	HIVE-19711 : Ferramenta Refactor Hive Schema
Backport	HIVE-19701 : getDelegationToken FromMetaStore não precisa ser sincronizado
Backport	HIVE-19694 : A instrução Create Materialized View deve verificar se há conflitos de nome da MV antes de executar a instrução da MV. SQL
Backport	HIVE-19674 : Agrupar por constantes decimais é empurrado para baixo até a tabela Druid
Backport	HIVE-19668 : Mais de 30% da pilha desperdiçada pela duplicação do org.antlr.runtime.CommonToken's e sequências duplicadas

Tipo	Descrição
Backport	HIVE-19663 : refatorar a geração de relatórios LLAP de IO
Backport	HIVE-19661 : altere o Hive UDFs para usar o motor regex Re2J
Backport	HIVE-19628 : possível em NPE LLAP testSigning
Backport	HIVE-19568 : HS2 HA ativo/passivo: não permitir conexão direta com a instância passiva HS2
Backport	HIVE-19564 : Vetorização: problemas de NULL correção/resultados errados em aritmética
Backport	HIVE-19552 : Habilite #druidkafkamini_basic .q TestMiniDruidKafkaCliDriver
Backport	HIVE-19432 : GetTablesOperation é muito lento se a colmeia tiver muitos bancos de dados e tabelas
Backport	HIVE-19360 :CBO: Adicione um “otimizadoSQL” ao objeto QueryPlan
Backport	HIVE-19326 : coleta automática de estatísticas: agregação incorreta durante as consultas UNION
Backport	HIVE-19313 : TestJdbcWith DBTokenStoreNoDoAs os testes estão falhando
Backport	HIVE-19285 : Adicione registros às subclasses de MetadataOperation

Tipo	Descrição
Backport	HIVE-19235 : Atualize arquivos dourados para testes Minimr
Backport	HIVE-19104 : Quando o teste MetaStore é iniciado com uma nova tentativa, as instâncias devem ser independentes
Backport	HIVE-18986 : A renomeação da tabela executará java.lang. StackOverflowError em dataNucleus se a tabela contiver um grande número de colunas
Backport	HIVE-18920 :CBO: Inicialize os provedores Janino antes da primeira consulta
Backport	HIVE-18873 : Ignorar a pressão silenciosa do predicado para MR em HiveInputFormat pode fazer com que os manipuladores de armazenamento produzam resultados errôneos
Backport	HIVE-18871 : erro de execução do hive on tez devido à configuração de hive.aux.jars.path como hdfs://
Backport	HIVE-18725 : Melhore o tratamento de erros para subconsultas se houver uma referência de coluna errada
Backport	HIVE-18696 : As pastas de partições podem não ser limpas adequadamente no HiveMetaStore método.add_partitions_core se um
Backport	HIVE-18453 :ACID: Adicione a sintaxe "CREATETRANSACTIONALTABLE" para unificar o suporte ao Parquet ACID ORC

Tipo	Descrição
Backport	HIVE-18201 : Desative XPROD _ EDGE para sq_count_chec
Backport	HIVE-18140 : Estatísticas de tabelas particionadas podem dar errado em estatísticas básicas em maiúsculas e minúsculas
Backport	HIVE-17921 : A agregação com struct in produz um resultado errado LLAP
Backport	HIVE-17896 : TopNKey: Crie um operador T vetorizável autônomo opNKey
Backport	HIVE-17840 : HiveMetaStore come exceção se. transactionalListeners notifyEvent falhar
Backport	HIVE-17043 : Remova colunas não exclusivas do grupo por chaves se não forem referenciadas posteriormente
Backport	HIVE-17040 : Eliminação conjunta na presença do relacionamento FK
Backport	HIVE-16839 : Chamadas desbalanceadas paraopenTransaction/commitTransaction ao alterar a mesma partição simultaneamente
Backport	HIVE-16100 : O otimizador dinâmico de partições classificadas perde operadores irmãos
Backport	HIVE-15956 : StackOverflowError quando soltar muitas partições

Tipo	Descrição
Backport	HIVE-15177 : A autenticação com o hive falha quando o tipo de autenticação kerberos é definido como e o principal contém _fromSubject HOST
Backport	HIVE-14898 : não HS2 deve registrar a pilha de chamadas devido a um erro de cabeçalho de autenticação vazio
Backport	HIVE-14493 : Suporte de particionamento para visualizações materializadas
Backport	HIVE-14431 : Reconhecer como COALESCE CASE
Backport	HIVE-13457 : Crie HS2 REST API endpoints para monitorar informações
Backport	HIVE-12342 : Defina o valor padrão de hive.optimize.index.filter como verdadeiro
Backport	HIVE-10296 : exceção de conversão observada quando o hive executa uma consulta de junção múltipla na metastore
Backport	HIVE-6980 : Elimine a tabela usando sql direto

Amazon EMR 6.6.0 - Alterações na configuração do Hive

- Como parte da OSS alteração [HIVE-20703](#), a propriedade para classificar partições dinâmicas, `hive.optimize.sort.dynamic.partition`, foi substituída por `hive.optimize.sort.dynamic.partition.threshold`

A configuração `hive.optimize.sort.dynamic.partition.threshold` tem os seguintes valores potenciais:

Value	Descrição
0 (padrão)	Torna a otimização para classificar partições dinâmicas uma decisão baseada em custos quando ORC os arquivos são usados. O número máximo de gravadores permitidos nas INSERT consultas é calculado com base em (memória do executor/contêiner) * (porcentagem da memória ocupada pelo orc) dividida pela memória máxima (tamanho da faixa) ocupada por um único gravador.
-1	Desabilita a otimização para classificar partições dinâmicas por completo.
1	Permite a classificação global de partições dinâmicas. Isso mantém apenas um gravador de registros aberto para cada valor de partição no redutor, reduzindo, portanto, a pressão de memória nos redutores.
2 (ou número inteiro maior)	Diz ao Hive para usar o número inteiro especificado como limite para o número máximo de gravadores.

Amazon EMR 6.6.0 - Problemas conhecidos do Hive

- Consultas com funções de janela na mesma coluna da junção podem levar a transformações inválidas, conforme relatado em [HIVE-25278](#), e causar resultados incorretos ou falhas na consulta. Como solução alternativa, você pode desativar essas consultas CBO no nível da consulta. Entre em contato com o AWS suporte para obter mais informações.
- A Amazon EMR 6.6.0 inclui a versão 3.1.2 do software Hive. [O Hive 3.1.2 introduz um recurso que divide arquivos de texto se eles contiverem cabeçalho e rodapé \(-21924\)](#). [HIVE](#) O Apache Tez App Master lê cada um dos seus arquivos para determinar pontos de deslocamento no intervalo de dados. Esses comportamentos combinados podem afetar negativamente a performance se suas consultas leem um grande número de pequenos arquivos de texto. Como solução alternativa, use `CombineHiveInputFormat` e ajuste o tamanho máximo da divisão configurando as seguintes propriedades:

```
SET hive.tez.input.format=org.apache.hadoop.hive.q1.io.CombineHiveInputFormat;
SET mapreduce.input.fileinputformat.split.maxsize=16777216;
```

- Com o Amazon EMR 6.6.0 a 6.9.x, INSERT as consultas com partição dinâmica e uma cláusula ORDER BY ou SORT BY sempre terão dois redutores. Esse problema é causado pela OSS alteração [HIVE-20703](#), que coloca a otimização dinâmica da partição de classificação sob decisão baseada em custos. Se sua workload não exigir a classificação de partições dinâmicas, recomendamos que você defina a propriedade `hive.optimize.sort.dynamic.partition.threshold` como -1 para desabilitar o novo atributo e obter o número de redutores calculado corretamente. Esse problema foi corrigido no OSS Hive como parte do [HIVE-22269](#) e foi corrigido no Amazon 6.10.0. EMR

Hudi

O [Apache Hudi](#) é uma estrutura de gerenciamento de dados de código aberto usada para simplificar o processamento incremental de dados e o desenvolvimento de pipelines de dados ao fornecer recursos de inserção, atualização, upsert e exclusão em nível de registro. Upsert refere-se à capacidade de inserir registros em um conjunto de dados existente se eles ainda não existirem ou atualizá-los se existirem. Ao gerenciar eficientemente a forma como os dados são apresentados no Amazon S3, o Hudi permite que os dados sejam ingeridos e atualizados quase em tempo real. O Hudi mantém cuidadosamente metadados das ações realizadas no conjunto de dados para ajudar a garantir que as ações sejam atômicas e consistentes.

O Hudi é integrado ao [Apache Spark](#), ao [Apache Hive](#) e ao [Presto](#). Nas versões 6.1.0 e posteriores da AmazonEMR, o Hudi também está integrado ao [Trino \(Presto\)](#). SQL

Com a EMR versão 5.28.0 e posterior da Amazon, EMR instala os componentes Hudi por padrão quando o Spark, o Hive, o Presto ou o Flink são instalados. Você pode usar o Spark ou o DeltaStreamer utilitário Hudi para criar ou atualizar conjuntos de dados Hudi. É possível usar o Hive, o Spark, o Presto ou o Flink para consultar um conjunto de dados do Hudi interativamente ou criar pipelines de processamento de dados usando extração incremental. Extração incremental refere-se à capacidade de extrair apenas os dados que foram alterados entre duas ações.

Esses atributos tornam o Hudi adequado para os seguintes casos de uso:

- Trabalhar com dados de streaming de sensores e outros dispositivos da Internet das Coisas (IoT) que exigem eventos específicos de inserção e atualização de dados.
- Cumprir os regulamentos de privacidade de dados em aplicativos em que os usuários possam optar por serem esquecidos ou modificar seu consentimento para a forma como os dados podem ser utilizados.
- Implementar um [sistema de captura de dados de alteração \(CDC\)](#) que permite aplicar alterações em um conjunto de dados ao longo do tempo.

A tabela a seguir lista a versão do Hudi incluída na versão mais recente da série Amazon EMR 7.x, junto com os componentes que a Amazon EMR instala com o Hudi.

Para a versão dos componentes instalados com o Hudi nesta versão, consulte [Versões de componentes da versão 7.2.0](#).

Informações sobre a versão Hudi para emr-7.2.0

Gravadora EMR de lançamento da Amazon	Versão do Hudi	Componentes instalados com o Hudi
emr-7.2.0	Hudi 0.14.1-amzn-1	Not available.

A tabela a seguir lista a versão do Hudi incluída na versão mais recente da série EMR 6.x da Amazon, junto com os componentes que a Amazon EMR instala com o Hudi.

Para obter a versão dos componentes instalados com o Hudi nessa versão, consulte [Release 6.15.0 Component Versions](#).

Informações de versão do Hudi para o emr-6.15.0

Gravadora EMR de lançamento da Amazon	Versão do Hudi	Componentes instalados com o Hudi
emr-6.15.0	Hudi 0.14.0-amzn-0	Not available.

Note

A EMR versão 6.8.0 da Amazon vem com o [Apache Hudi](#) 0.11.1; no entanto, os clusters da EMR Amazon 6.8.0 também são compatíveis com o código aberto do Hudi 0.12.0. `hudi-spark3.3-bundle_2.12`

A tabela a seguir lista a versão do Hudi incluída na versão mais recente da série Amazon EMR 5.x, junto com os componentes que a Amazon EMR instala com o Hudi.

Para a versão dos componentes instalados com o Hudi nesta versão, consulte [Versões de componentes da versão 5.36.2](#).

Informações sobre a versão Hudi para emr-5.36.2

Gravadora EMR de lançamento da Amazon	Versão do Hudi	Componentes instalados com o Hudi
emr-5.36.2	Hudi 0.10.1-amzn-1	Not available.

Tópicos

- [Como o Hudi funciona](#)
- [Considerações e limitações para usar o Hudi na Amazon EMR](#)
- [Criar um cluster com o Hudi instalado](#)
- [Trabalhe com um conjunto de dados do Hudi](#)
- [Use o Hudi CLI](#)
- [Histórico de versões do Hudi](#)

Como o Hudi funciona

Ao usar o Hudi com a AmazonEMR, você pode gravar dados no conjunto de dados usando a fonte de dados Spark API ou o utilitário Hudi. DeltaStreamer O Hudi organiza um conjunto de dados em uma estrutura de diretórios particionada em um *basepath* que é semelhante a uma tabela tradicional do Hive. As especificidades de como os dados são definidos como arquivos nesses diretórios dependem do tipo de conjunto de dados escolhido. É possível escolher Copiar na gravação (CoW – Copy on Write) ou Mesclar na leitura (MoR – Merge on Read).

Independentemente do tipo de conjunto de dados, cada partição em um conjunto de dados é identificada exclusivamente por seu *partitionpath* relativo ao *basepath*. Dentro de cada partição, os registros são distribuídos em vários arquivos de dados. Para obter mais informações, consulte [File management](#) na documentação do Apache Hudi.

Cada ação no Hudi tem uma confirmação correspondente, identificada por um carimbo de data/hora monotonicamente crescente conhecido como Instant. O Hudi mantém uma série de todas as ações executadas no conjunto de dados como uma linha do tempo. O Hudi conta com a linha do tempo para fornecer isolamento de snapshot entre leitores e gravadores e para permitir a reversão para um momento anterior específico. Para obter mais informações sobre as ações registradas pelo Hudi e o estado das ações, consulte [Linha do tempo](#) na documentação do Apache Hudi.

Noções básicas sobre os tipos de armazenamento de conjuntos de dados: copiar na gravação ou mesclar na leitura

Ao criar um conjunto de dados do Hudi, você especifica se o conjunto de dados é do tipo copiar na gravação ou mesclar na leitura.

- **Copy on Write (CoW – Copiar na gravação):** os dados são armazenados em um formato colunar (Parquet), e cada atualização cria uma nova versão dos arquivos durante uma gravação. CoW é o tipo de armazenamento padrão.
- **Merge on Read (MoR – Mesclar na leitura):** os dados são armazenados usando uma combinação de formatos colunares (Parquet) e baseados em linha (Avro). As atualizações são registradas em arquivos delta baseados em linha e são compactadas conforme necessário para criar novas versões dos arquivos colunares.

Com conjuntos de dados CoW, sempre que há uma atualização para um registro, o arquivo que contém o registro é regravado com os valores atualizados. Com um conjunto de dados MoR, sempre que há uma atualização, o Hudi grava apenas a linha do registro alterado. MoR é mais adequado para cargas de trabalho com maior volume de gravações ou alterações e menor volume de leituras. CoW é mais adequado para cargas de trabalho com maior volume de leituras em dados que mudam com menos frequência.

O Hudi fornece três visualizações lógicas para acessar os dados:

- **Visão otimizada para leitura:** fornece o conjunto de dados confirmado mais recente de tabelas CoW e o conjunto de dados compactados mais recente de tabelas MoR.
- **Visualização incremental** — fornece um fluxo de mudança entre duas ações de um conjunto de dados CoW para alimentar trabalhos posteriores e fluxos de trabalho de extração, transformação e carregamento (ETL).
- **Visão em tempo real:** fornece os dados confirmados mais recentes de uma tabela MoR ao mesclar os arquivos baseados em colunas e linhas em linha.

Quando você consulta a visualização otimizada para leitura, a consulta retorna todos os dados compactados, mas não inclui as últimas confirmações delta. A consulta desses dados apresenta bom desempenho de leitura, mas omite os dados mais recentes. Quando você consulta a visualização em tempo real, o Hudi mescla os dados compactados com as confirmações delta na leitura. Os dados mais recentes estão disponíveis para consulta, mas a sobrecarga computacional da mesclagem prejudica o desempenho da consulta. A capacidade de consultar dados compactados ou dados em tempo real permite a você escolher entre o desempenho e a flexibilidade nas consultas.

Para obter mais informações sobre as vantagens e desvantagens dos tipos de armazenamento, consulte [Tipos e visões de armazenamento](#) na documentação do Apache Hudi.

O Hudi cria duas tabelas no metastore do Hive para MoR: uma tabela com o nome especificado por você, que é uma visão otimizada para leitura, e uma tabela com o mesmo nome anexado com `_rt`, que é uma visão em tempo real. É possível consultar as duas tabelas.

Registrar um conjunto de dados do Hudi com seu metastore

Ao registrar uma tabela Hudi na metastore Hive, você pode consultar tabelas Hudi usando Hive, Spark SQL ou Presto como faria com qualquer outra tabela. Além disso, você pode integrar o Hudi com o AWS Glue configurando o Hive e o Spark para usar o AWS Glue Data Catalog como metastore. Para tabelas MoR, o Hudi registra o conjunto de dados como duas tabelas no Metastore: uma tabela com o nome especificado por você, que é uma visão otimizada para leitura, e uma tabela com o mesmo nome anexado com `_rt`, que é uma visão em tempo real.

Você registra uma tabela do Hudi com o metastore do Hive ao usar o Spark para criar um conjunto de dados do Hudi ao definir a opção `HIVE_SYNC_ENABLED_OPT_KEY` como `"true"` e fornecendo outras propriedades necessárias. Para ter mais informações, consulte [Trabalhe com um conjunto de dados do Hudi](#). Além disso, é possível usar o utilitário de linha de comando `hive_sync_tool` para registrar um conjunto de dados do Hudi como uma tabela no seu metastore, separadamente.

Considerações e limitações para usar o Hudi na Amazon EMR

- O campo de chave de registro não pode ser nulo nem estar em branco: o campo especificado como o campo de chave de registro não pode ter valores `null` nem em branco.
- Esquema atualizado por padrão ao upsert e insert — o Hudi fornece uma interface `HoodieRecordPayload` que determina como a entrada `DataFrame` e o conjunto de dados Hudi existente são mesclados para produzir um conjunto de dados novo e atualizado. O Hudi fornece uma implementação padrão dessa classe, `OverwriteWithLatestAvroPayload`, que sobrescreve os registros existentes e atualiza o esquema conforme especificado na entrada `DataFrame`. Para personalizar essa lógica a fim de implementar atualizações parciais e de mesclagem, é possível fornecer uma implementação da interface `HoodieRecordPayload` usando o parâmetro `DataSourceWriteOptions.PAYLOAD_CLASS_OPT_KEY`.
- A exclusão requer um esquema: ao excluir, você precisa especificar a chave de registro, a chave de partição e os campos de chave de pré-combinação. Outras colunas podem ficar `null` ou em branco, mas o esquema completo é necessário.
- Limitações das tabelas MoR: as tabelas MoR não são compatíveis com pontos de salvamento. Você pode consultar tabelas MoR usando a visualização otimizada para leitura ou a visualização

em tempo real (*tableName_rt*) do SparkSQL, Presto ou Hive. O uso da visão otimizada para leitura expõe apenas os dados do arquivo básico e não expõe uma visão mesclada dos dados básicos e de log.

- Hive

- Para registrar tabelas no metastore do Hive, o Hudi espera que o servidor Thrift do Hive esteja em execução na porta padrão 10000. Se você substituir essa porta por uma porta personalizada, transmita a opção HIVE_URL_OPT_KEY conforme mostrado no exemplo a seguir.

```
.option(DataSourceWriteOptions.HIVE_URL_OPT_KEY, "jdbc:hive2://localhost:override-port-number
```

- O tipo de dados timestamp no Spark é registrado como tipo de dados long no Hive, e não como o tipo timestamp do Hive.

- Presto

- O Presto não é compatível com a leitura de tabelas MoR em tempo real em versões do Hudi anteriores à 0.6.0.
- O Presto só oferece suporte a consultas de snapshots.
- Para o Presto interpretar corretamente colunas de conjuntos de dados do Hudi, defina o valor `hive.parquet_use_column_names` como `true`.
 - Para definir o valor de uma sessão, execute o seguinte comando no shell Presto:

```
set session hive.parquet_use_column_names=true
```

- Para definir o valor no nível do cluster, use a classificação de configuração `presto-connector-hive` para definir `hive.parquet.use_column_names` como `true`, conforme mostrado no exemplo a seguir. Para ter mais informações, consulte [Configurar aplicações](#).

```
[  
  {  
    "Classification": "presto-connector-hive",  
    "Properties": {  
      "hive.parquet.use-column-names": "true"  
    }  
  }  
]
```

- HBaseÍndice

- A HBase versão usada para criar o Hudi pode ser diferente da listada no Guia de EMR Lançamento. Para extrair as dependências corretas da sua sessão do Spark, acesse o comando a seguir.

```
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"
```

Criar um cluster com o Hudi instalado

Com a EMR versão 5.28.0 e posterior da Amazon, a Amazon EMR instala componentes Hudi por padrão quando o Spark, o Hive ou o Presto são instalados. Para usar o Hudi na AmazonEMR, crie um cluster com um ou mais dos seguintes aplicativos instalados:

- Hadoop
- Hive
- Spark
- Presto
- Flink

Você pode criar um cluster usando o AWS Management Console AWS CLI, o ou o Amazon EMR API.

Para criar um cluster com o Hudi usando o AWS Management Console

1. Navegue até o novo EMR console da Amazon e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte [Usar o console antigo](#).
2. Escolha Create cluster (Criar cluster), Go to advanced options (Ir para opções avançadas).
3. Em Software Configuration (Configuração de software), escolha emr-5.28.0 ou posterior para Release (Versão) e selecione Hadoop, Hive, Spark, Presto e Tez junto com outros aplicativos exigidos pelo seu cluster.
4. Configure outras opções conforme necessário para seu aplicativo e escolha Next (Avançar).
5. Configure as opções de Hardware e as General cluster settings (Configurações gerais do cluster) conforme desejado.

6. Para Opções de segurança, recomendamos que você selecione um par de EC2 chaves que possa ser usado para se conectar à linha de comando do nó principalSSH. Isso permite que você execute os comandos shell do Spark, os comandos do Hive e CLI os comandos do Hudi CLI descritos neste guia.
7. Configure outras opções de segurança conforme desejado e escolha Create cluster (Criar cluster).

Trabalhe com um conjunto de dados do Hudi

O Hudi oferece suporte à inserção, atualização e exclusão de dados em conjuntos de dados do Hudi por meio do Spark. Para obter mais informações, consulte [Gravar tabelas do Hudi](#) na documentação do Apache Hudi.

Os exemplos a seguir demonstram como iniciar o shell interativo do Spark, usar o envio do Spark ou usar o Amazon EMR Notebooks para trabalhar com o Hudi na Amazon. EMR Você também pode usar o DeltaStreamer utilitário Hudi ou outras ferramentas para gravar em um conjunto de dados. Nesta seção, os exemplos demonstram como trabalhar com conjuntos de dados usando o shell do Spark enquanto estão conectados ao nó principal usando SSH como usuário padrãohadoop.

Inicie o shell Spark usando o Amazon EMR 6.7 e versões posteriores

Ao executar `spark-shell` ou `spark-sql` usar o Amazon EMR 6.7.0 ou posterior, passe os seguintes comandos. `spark-submit`

Note

A Amazon EMR 6.7.0 usa o [Apache Hudi 0.11.0-amzn-0](#), que contém melhorias significativas em relação às versões anteriores do Hudi. Para obter mais informações, consulte o [Guia de migração do Apache Hudi 0.11.0](#). Os exemplos nesta guia refletem essas alterações.

Abrir o shell do Spark no nó primário

1. Conecte-se ao nó primário usando SSH o. Para obter mais informações, consulte [Conecte-se ao nó primário usando SSH](#) o Amazon EMR Management Guide.
2. Digite o seguinte comando para iniciar o shell do Spark. Para usar a PySpark concha, substitua `spark-shell` with `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"
```

Inicie o shell Spark usando o Amazon EMR 6.6 e versões anteriores

Ao executar `spark-shell` ou `spark-sql` usar o Amazon EMR 6.6.x ou anterior, passe os seguintes comandos. `spark-submit`

Note

- O Amazon EMR 6.2 e 5.31 e versões posteriores (Hudi 0.6.x e versões posteriores) podem omitir o da configuração. `spark-avro.jar`
- O Amazon EMR 6.5 e 5.35 e versões posteriores (Hudi 0.9.x e versões posteriores) podem ser omitidos da configuração. `spark.sql.hive.convertMetastoreParquet=false`
- [O Amazon EMR 6.6 e 5.36 e versões posteriores \(Hudi 0.10.x e versões posteriores\) devem incluir a HoodieSparkSessionExtension configuração conforme descrito na versão: 0.10.0 Spark Guide:](#)

```
--conf  
"spark.sql.extensions=org.apache.spark.sql.hudi.HoodieSparkSessionExtension"  
\  

```

Abrir o shell do Spark no nó primário

1. Conecte-se ao nó primário usando SSH o. Para obter mais informações, consulte [Conecte-se ao nó primário usando SSH](#) o Amazon EMR Management Guide.
2. Digite o seguinte comando para iniciar o shell do Spark. Para usar a PySpark concha, substitua `spark-shell` with `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
```

Use o Hudi com Amazon EMR Notebooks usando o Amazon EMR 6.7 e versões posteriores

Para usar o Hudi com o Amazon EMR Notebooks, você deve primeiro copiar os arquivos jar do Hudi do sistema de arquivos local para HDFS o nó principal do cluster de notebooks. Em seguida, você usa o editor do notebook para configurar seu EMR notebook para usar o Hudi.

Para usar o Hudi com Amazon Notebooks EMR

1. Crie e lance um cluster para Amazon EMR Notebooks. Para obter mais informações, consulte [Criação de EMR clusters da Amazon para notebooks](#) no Amazon EMR Management Guide.
2. Conecte-se ao nó principal do cluster usando SSH e, em seguida, copie os arquivos jar do sistema de arquivos local para, HDFS conforme mostrado nos exemplos a seguir. No exemplo, criamos um diretório HDFS para facilitar o gerenciamento de arquivos. Você pode escolher seu próprio destino em HDFS, se desejar.

```
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. Abra o editor do notebook, insira o código do exemplo a seguir e execute-o.

```
%%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"  
}}
```

Use o Hudi com Amazon EMR Notebooks usando o Amazon EMR 6.6 e versões anteriores

Para usar o Hudi com o Amazon EMR Notebooks, você deve primeiro copiar os arquivos jar do Hudi do sistema de arquivos local para HDFS o nó principal do cluster de notebooks. Em seguida, você usa o editor do notebook para configurar seu EMR notebook para usar o Hudi.

Para usar o Hudi com Amazon Notebooks EMR

1. Crie e lance um cluster para Amazon EMR Notebooks. Para obter mais informações, consulte [Criação de EMR clusters da Amazon para notebooks](#) no Amazon EMR Management Guide.
2. Conecte-se ao nó principal do cluster usando SSH e, em seguida, copie os arquivos jar do sistema de arquivos local para, HDFS conforme mostrado nos exemplos a seguir. No exemplo, criamos um diretório HDFS para facilitar o gerenciamento de arquivos. Você pode escolher seu próprio destino em HDFS, se desejar.

```
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. Abra o editor do notebook, insira o código do exemplo a seguir e execute-o.

```
{ "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"  
}}
```

Inicializar uma sessão do Spark para Hudi

Ao usar o Scala, você deve importar as seguintes classes na sessão do Spark. Isso precisa ser feito uma vez por sessão do 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
```

Gravar em um conjunto de dados do Hudi

Os exemplos a seguir mostram como criar um DataFrame e escrevê-lo como um conjunto de dados Hudi.

Note

Para colar exemplos de código no shell do Spark, digite **:paste** no prompt, cole o exemplo e pressione **CTRL + D**.

Cada vez que você grava um DataFrame em um conjunto de dados Hudi, você deve especificar. `DataSourceWriteOptions` Muitas dessas opções provavelmente serão idênticas entre as operações de gravação. O exemplo a seguir especifica opções comuns usando a variável `hudiOptions`, usada pelos exemplos subsequentes.

Escreva usando Scala com o Amazon EMR 6.7 e versões posteriores

Note

A Amazon EMR 6.7.0 usa o [Apache Hudi 0.11.0-amzn-0](#), que contém melhorias significativas em relação às versões anteriores do Hudi. Para obter mais informações, consulte o [Guia de migração do Apache Hudi 0.11.0](#). Os exemplos nesta guia refletem essas alterações.

```
// 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.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.MultiPartKeysValueExtractor",
  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/"))

```

Escreva usando Scala com o Amazon EMR 6.6 e versões anteriores

```

// 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/"))
```

Escreva usando 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.MultiPartKeyValueExtractor'
}

# 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

Você pode ver “hoodie” em vez de Hudi em exemplos de código e notificações. A base de código do Hudi usa amplamente a antiga grafia “hoodie”.

DataSourceWriteOptions referência para Hudi

Opção	Descrição
TABLE_NAME	O nome da tabela com o qual registrar o conjunto de dados.
TABLE_TYPE_OPT_KEY	Opcional. Especifica se o conjunto de dados foi criado como "COPY_ON_WRITE" ou "MERGE_ON_READ" . O padrão é "COPY_ON_WRITE" .
RECORDKEY_FIELD_OPT_KEY	O campo de chave de registro cujo valor será usado como o componente recordKey de HoodieKey . O valor real será obtido invocando .toString() no valor do campo. Campos aninhados podem ser especificados

Opção	Descrição
	usando a notação de pontos, por exemplo, <code>a.b.c</code> .
<code>PARTITIONPATH_FIELD_OPT_KEY</code>	O campo de caminho de partição cujo valor será usado como o componente <code>partition Path</code> de <code>HoodieKey</code> . O valor real será obtido invocando <code>.toString()</code> no valor do campo.
<code>PRECOMBINE_FIELD_OPT_KEY</code>	O campo usado na pré-combinação antes da gravação real. Quando dois registros têm o mesmo valor de chave, o Hudi seleciona aquele com o maior valor para o campo de pré-combinação, conforme determinado por <code>Object.compareTo(..)</code> .

As opções a seguir são necessárias apenas para registrar a tabela do conjunto de dados do Hudi no seu metastore. Se você não registrar o conjunto de dados do Hudi como uma tabela no metastore do Hive, essas opções não serão necessárias.

`DataSourceWriteOptions` referência para Hive

Opção	Descrição
<code>HIVE_DATABASE_OPT_KEY</code>	O banco de dados do Hive com o qual sincronizar. O padrão é <code>"default"</code> .
<code>HIVE_PARTITION_EXTRACTOR_CLASS_OPT_KEY</code>	A classe usada para extrair valores de campo de partição em colunas de partição do Hive.
<code>HIVE_PARTITION_FIELDS_OPT_KEY</code>	O campo no conjunto de dados a ser usado para determinar colunas de partição do Hive.
<code>HIVE_SYNC_ENABLED_OPT_KEY</code>	Quando definido como <code>"true"</code> , registra o conjunto de dados no metastore do Apache Hive. O padrão é <code>"false"</code> .

Opção	Descrição
HIVE_TABLE_OPT_KEY	Obrigatório. O nome da tabela no Hive com a qual sincronizar. Por exemplo, "my_hudi_table_cow" .
HIVE_USER_OPT_KEY	Opcional. O nome de usuário do Hive a ser usado ao sincronizar. Por exemplo, "hadoop".
HIVE_PASS_OPT_KEY	Opcional. A senha do Hive para o usuário especificado por HIVE_USER_OPT_KEY .
HIVE_URL_OPT_KEY	A URL metastore Hive.

Upsert dados

O exemplo a seguir demonstra como alterar dados escrevendo um DataFrame Ao contrário do exemplo anterior de inserção, o valor OPERATION_OPT_KEY é definido como UPSERT_OPERATION_OPT_VAL. Além disso, .mode(SaveMode.Append) é especificado para indicar que o registro deve ser anexado.

Upsert usando Scala com Amazon EMR 6.7 e versões posteriores

Note

A Amazon EMR 6.7.0 usa o [Apache Hudi 0.11.0-amzn-0](#), que contém melhorias significativas em relação às versões anteriores do Hudi. Para obter mais informações, consulte o [Guia de migração do Apache Hudi 0.11.0](#). Os exemplos nesta guia refletem essas alterações.

```
// 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 usando Scala com Amazon EMR 6.6 e versões anteriores

```
// 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/"))
```

Interromper usando 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(**hudiOptions) \
  .mode('append') \
  .save('s3://DOC-EXAMPLE-BUCKET/myhuidataset/')
```

Excluir um registro

Para excluir um registro de forma irreversível, você pode upsert uma carga útil vazia. Nesse caso, a opção `PAYLOAD_CLASS_OPT_KEY` especifica a classe `EmptyHoodieRecordPayload`. O exemplo usa o mesmo `DataFrame`, `updateDF`, usado no exemplo upsert para especificar o mesmo registro.

Exclua usando o Scala com o Amazon EMR 6.7 e versões posteriores

Note

A Amazon EMR 6.7.0 usa o [Apache Hudi 0.11.0-amzn-0](#), que contém melhorias significativas em relação às versões anteriores do Hudi. Para obter mais informações, consulte o [Guia de migração do Apache Hudi 0.11.0](#). Os exemplos nesta guia refletem essas alterações.

```
(updateDF.write
  .format("hudi")
  .options(hudiOptions)
  .option(DataSourceWriteOptions.OPERATION_OPT_KEY, "delete")
  .mode(SaveMode.Append)
  .save("s3://DOC-EXAMPLE-BUCKET/myhuidataset/"))
```

Exclua usando o Scala com o Amazon EMR 6.6 e versões anteriores

```
(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/myhuidataset/"))
```

Excluir usando 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/myhuidataset/')
```

Você também pode excluir dados de forma irreversível definindo `OPERATION_OPT_KEY` como `DELETE_OPERATION_OPT_VAL` para remover todos os registros no conjunto de dados enviado.

Para obter instruções sobre como realizar exclusões reversíveis e obter mais informações sobre a exclusão de dados armazenados em tabelas do Hudi, consulte [Exclusões](#) na documentação do Apache Hudi.

Ler em um conjunto de dados do Hudi

Para recuperar dados no momento atual, o Hudi realiza consultas de snapshots por padrão. Veja a seguir um exemplo de consulta do conjunto de dados gravado no S3 em [Gravar em um conjunto de dados do Hudi](#). Substituir `s3://DOC-EXAMPLE-BUCKET/myhudidataset` com o caminho da tabela e adicione asteriscos curinga para cada nível de partição, além de um asterisco adicional. Neste exemplo, há um nível de partição, portanto adicionamos dois símbolos curinga.

Leia usando o Scala com o Amazon EMR 6.7 e versões posteriores

Note

A Amazon EMR 6.7.0 usa o [Apache Hudi](#) 0.11.0-amzn-0, que contém melhorias significativas em relação às versões anteriores do Hudi. Para obter mais informações, consulte o [Guia de migração do Apache Hudi 0.11.0](#). Os exemplos nesta guia refletem essas alterações.

```
val snapshotQueryDF = spark.read
  .format("hudi")
  .load("s3://DOC-EXAMPLE-BUCKET/myhudidataset")
  .show()
```

Leia usando o Scala com o Amazon EMR 6.6 e versões anteriores

```
(val snapshotQueryDF = spark.read
  .format("org.apache.hudi")
  .load("s3://DOC-EXAMPLE-BUCKET/myhudidataset" + "/*/*"))

snapshotQueryDF.show()
```

Leia usando PySpark

```
snapshotQueryDF = spark.read \
  .format('org.apache.hudi') \
  .load('s3://DOC-EXAMPLE-BUCKET/myhudidataset' + '/*/*')
```

```
snapshotQueryDF.show()
```

Consultas incrementais

Você também pode realizar consultas incrementais com o Hudi para obter um fluxo de registros que foram alterados desde um determinado carimbo de data/hora de confirmação. Para fazer isso, defina o campo `QUERY_TYPE_OPT_KEY` como `QUERY_TYPE_INCREMENTAL_OPT_VAL`. Em seguida, adicione um valor para `BEGIN_INSTANTTIME_OPT_KEY` para obter todos os registros gravados desde a hora especificada. Normalmente, as consultas incrementais são dez vezes mais eficientes do que as de lote, pois processam somente registros alterados.

Ao realizar consultas incrementais, use o caminho da tabela raiz (básica) sem os asteriscos curinga usados nas consultas Snapshot.

Note

O Presto não é compatível com consultas incrementais.

Consultas incrementais usando o 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()
```

Consultas incrementais usando 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/myhuidataset')  
  
incQueryDF.show()
```

Para obter mais informações sobre a leitura de conjuntos de dados do Hudi, consulte [Consultar tabelas do Hudi](#) na documentação do Apache Hudi.

Use o Hudi CLI

Você pode usar o Hudi CLI para administrar conjuntos de dados Hudi para visualizar informações sobre confirmações, sistema de arquivos, estatísticas e muito mais. Você também pode usar o CLI para realizar compactações manualmente, programar compactações ou cancelar compactações programadas. Para obter mais informações, consulte [Interagindo via CLI](#) na documentação do Apache Hudi.

Para iniciar o Hudi CLI e conectar-se a um conjunto de dados

1. Conecte-se ao nó principal usando SSH o. Para obter mais informações, consulte [Conecte-se ao nó principal usando SSH](#) o Amazon EMR Management Guide.
2. Na linha de comando, digite `/usr/lib/hudi/cli/bin/hudi-cli.sh`.

O prompt de comando muda para `hudi->`.

3. Digite o seguinte código para se conectar a um conjunto de dados. Substituir `s3://DOC-EXAMPLE-BUCKET/myhuidataset` com o caminho para o conjunto de dados com o qual você deseja trabalhar. O valor que usamos é o mesmo estabelecido em exemplos anteriores.

```
connect --path s3://DOC-EXAMPLE-BUCKET/myhuidataset
```

O prompt de comando é alterado para incluir o conjunto de dados ao qual você está conectado, conforme mostrado no exemplo a seguir.

```
hudi:myhuidataset->
```

Histórico de versões do Hudi

A tabela a seguir lista a versão do Hudi incluída em cada versão de lançamento da AmazonEMR, junto com os componentes instalados com o aplicativo. Para obter as versões dos componentes em

cada versão, consulte a seção Versão de componentes da sua versão em [Versões de lançamento do Amazon EMR 7.x](#), [Versões de lançamento do Amazon EMR 6.x](#) ou [Versões de lançamento do Amazon EMR 5.x](#).

Informações de versão do Hudi

Gravadora Amazon EMR Release	Versão do Hudi	Componentes instalados com o 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.
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.

Gravadora Amazon EMR Release	Versão do Hudi	Componentes instalados com o Hudi
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.
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.

Gravadora Amazon EMR Release	Versão do Hudi	Componentes instalados com o Hudi
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.
emr-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

O Hue (Hadoop User Experience) é uma interface gráfica de usuário de código aberto, baseada na Web, para uso com a Amazon e o Apache Hadoop. O Hue agrupa vários projetos de ecossistemas do Hadoop diferentes em uma interface configurável. A Amazon também adicionou personalizações específicas para o Hue na Amazon EMR. O Hue atua como um front-end para aplicativos executados no cluster, permitindo que você interaja com os aplicativos usando uma interface que possa ser mais familiar ou amigável para o usuário. Os aplicativos no Hue, como os editores do Hive e do Pig, dispensam a necessidade de fazer login no cluster para executar scripts interativamente usando o shell de cada aplicativo. Depois que um cluster for iniciado, você pode interagir totalmente com os aplicativos usando o Hue ou uma interface similar. Para obter mais informações sobre o Hue, consulte <http://gethue.com>.

O Hue é instalado por padrão quando você executa seu cluster usando o EMR console da Amazon. Você pode optar por não instalar o Hue usando as opções avançadas no EMR console da Amazon ao iniciar um cluster ou especificando explicitamente a `--applications` opção e omitindo o Hue ao usar o `create-cluster` AWS CLI

Tópicos

- [Informações de versão do Hue](#)
- [Recursos compatíveis e não suportados do Hue na Amazon EMR](#)
- [Considerações](#)
- [Conectar-se à interface de usuário da Web do Hue](#)
- [Usando o Hue com um banco de dados remoto na Amazon RDS](#)
- [Configurações avançadas do Hue](#)
- [Histórico de versões do Hue](#)

Informações de versão do Hue

Versão Hue para 7.2.0

A tabela a seguir lista a versão do Hue incluída na versão mais recente da série Amazon EMR 7.x, junto com os componentes que a Amazon EMR instala com o Hue.

Para a versão dos componentes instalados com o Hue nesta versão, consulte [Versões de componentes da versão 7.2.0](#).

Informações sobre a versão do Hue para o emr-7.2.0

Gravadora EMR de lançamento da Amazon	Versão do Hue	Componentes instalados com o 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

Versão do Hue para 6.15.0

A tabela a seguir lista a versão do Hue incluída na versão mais recente da série EMR 6.x da Amazon, junto com os componentes que a Amazon EMR instala com o Hue.

Para obter a versão dos componentes instalados com o Hue nessa versão, consulte [Release 6.15.0 Component Versions](#).

Informações de versão do Hue para o emr-6.15.0

Gravadora EMR de lançamento da Amazon	Versão do Hue	Componentes instalados com o 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-

Gravadora EMR de lançamento da Amazon	Versão do Hue	Componentes instalados com o Hue
		timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Versão Hue para 5.36.2

A tabela a seguir lista a versão do Hue incluída na versão mais recente da série Amazon EMR 5.x, junto com os componentes que a Amazon EMR instala com o Hue.

Para a versão dos componentes instalados com o Hue nesta versão, consulte Versões de componentes da [versão 5.36.2](#).

Informações sobre a versão do Hue para o emr-5.36.2

Gravadora EMR de lançamento da Amazon	Versão do Hue	Componentes instalados com o 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-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Recursos compatíveis e não suportados do Hue na Amazon EMR

Com o Amazon EMR 7.0.0 e versões posteriores, o Hue exige o Python 3.9 ou superior. Ao usar criptografia em trânsito, é necessário usar uma chave de cifra de 2048 bits.

- Amazon S3 e Hadoop File System (HDFS) Browser

- Com as permissões apropriadas, você pode navegar e mover dados entre o HDFS armazenamento temporário e os buckets S3 pertencentes à sua conta.
- Por padrão, os superusuários no Hue podem acessar todos os arquivos que as EMR IAM funções da Amazon têm permissão para acessar. Os usuários recém-criados não têm automaticamente permissões para acessar o navegador de arquivos Amazon S3 e devem ter as permissões `filebrowser.s3_access` ativadas para o grupo deles.
- Hive: execute consultas interativas nos seus dados. Esta também é uma maneira útil de gerar protótipos de consultas programáticas ou em lotes.
- Pig: execute scripts nos seus dados ou emita comandos interativos.
- Oozie: crie e monitore fluxos de trabalho do Oozie.
- Metastore Manager: visualize e manipule o conteúdo do Hive Metastore (importar, criar, descartar e assim por diante).
- Navegador de trabalhos: veja o status dos seus trabalhos do Hadoop enviados.
- Gerenciamento de usuários — Gerencie contas do Hue e integre LDAP usuários ao Hue.
- AWS Amostras — Há vários exemplos ready-to-run "" que processam dados de amostra de vários AWS serviços usando aplicativos no Hue. Ao fazer logon no Hue, você será levado ao aplicativo Hue Home, onde os exemplos estão pré-instalados.
- O Livy Server é suportado somente na Amazon EMR versão 5.9.0 e posterior.
- Para usar o Hue Notebook para Spark, você deve instalar o Hue com Livy e Spark.
- Trino/Presto: o Hue oferece suporte à conexão com os servidores de dados Trino e Presto. Com o Amazon EMR 7.0.0 e versões posteriores, essa conectividade é facilitada com `trino-python-client` conectores. `presto-python-client` Versões anteriores à 7.0.0 usam um JDBC conector por meio de um JDBC gateway.
- Não há suporte para o painel do Hue.
- O Postgre não SQL é suportado.

Considerações

Considere as seguintes limitações ao usar o [Hue](#) na AmazonEMR.

Desempenho com grandes tabelas de metadados do Hue

Se o banco de dados de metadados do Hue ficar muito grande, o desempenho poderá diminuir. Para verificar o tamanho das tabelas, primeiro conecte-se ao nó primário do EC2 cluster Amazon EMR on

Amazon SSH e execute o comando `sudo mysql -u root` para iniciar o My SQLCLI. Para obter o tamanho da sua tabela, execute a consulta `SELECT COUNT(*) FROM hue.<table_name>`. Veja o seguinte para saber o que `<table_name>` pode ser:

- documento_de_desktop
- documento_de_área de trabalho 2
- oozie_job
- consulta beeswax_saved
- sessão de cera de abelha
- histórico de query_beeswax_

Se a execução dessa consulta retornar uma contagem de mais de 100000, você deverá executar o seguinte comando de limpeza para excluir os registros antigos.

```
cd /opt/cloudera/parcels/CDH/lib/hue # Hue home directory ./build/env/bin/hue
desktop_document_cleanup
```

Para obter mais informações sobre como limpar seu banco de dados, consulte a [página de referência na documentação do Hue](#).

O Hue não limpa automaticamente as tabelas, mas as EMR versões 5.12.0 e superiores da Amazon fornecem um método para excluir documentos antigos nas tabelas. Crie o script de shell a seguir e execute-o como uma etapa em um EMR cluster da Amazon com um parâmetro inteiro que representa o número máximo de dias de quanto tempo manter documentos no banco de dados de metadados.

```
#!/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
```

Incompatibilidade entre as versões do Hue

Se você estiver usando o mesmo banco de dados de metadados do Hue em vários clusters habilitados para o Hue, recomendamos que esses clusters executem a mesma versão do Hue. Versões diferentes do Hue podem ter esquemas diferentes para o banco de dados de metadados do Hue. Usar o mesmo banco de dados para versões diferentes pode causar falha na instalação do Hue. Por exemplo, tentar usar o mesmo banco de dados para dois clusters com 4.10.0 e 4.11.0 instalados pode causar erros de login para os usuários que estão tentando fazer login no cluster 4.10.0.

Para obter uma lista das etiquetas de EMR lançamento da Amazon e das versões instaladas correspondentes do Hue, consulte o histórico de [lançamentos do Hue](#).

Conectar-se à interface de usuário da Web do Hue

Conectar-se à interface de usuário da web do Hue é o mesmo que conectar-se a qualquer HTTP interface hospedada no nó principal de um cluster. O procedimento a seguir descreve como acessar a interface de usuário do Hue. Para obter mais informações, consulte [Visualizar interfaces web hospedadas em EMR clusters](#) no Amazon EMR Management Guide.

Visualizar a interface de usuário da Web do Hue

1. Siga estas instruções para [configurar um SSH túnel para o nó principal usando o encaminhamento dinâmico de portas](#) no Amazon EMR Management Guide.
2. Digite o seguinte endereço no seu navegador para abrir a interface web do Hue: onde `http://master public DNS:8888master public dns` é o DNS nome público do nó principal do cluster, por exemplo `ec2-11-22-333-44.compute-1.amazonaws.com`.
3. Na tela de login do Hue, se você for o administrador fazendo login pela primeira vez, insira um nome de usuário e uma senha para criar sua superconta no Hue e selecione Criar conta. Caso contrário, digite seu nome de usuário e sua senha e selecione Create account (Criar conta) ou insira as credenciais fornecidas pelo administrador.

Usando o Hue com um banco de dados remoto na Amazon RDS

Por padrão, as informações do usuário e os históricos de consultas do Hue são armazenados em um SQL banco de dados My local no nó principal. Como alternativa, você pode criar um ou mais clusters habilitados para o HUE usando uma configuração armazenada no Amazon S3 e um banco de dados SQL My no Amazon Relational Database Service (Amazon). RDS Isso permite que você mantenha as informações do usuário e o histórico de consultas criado pelo Hue sem manter seu EMR cluster da Amazon em execução. Recomendamos o uso da criptografia do lado do servidor do Amazon S3 para armazenar o arquivo de configuração.


Primeiro, crie o banco de dados remoto para o Hue.

Para criar o Meu SQL banco de dados externo

1. Abra o RDS console da Amazon em <https://console.aws.amazon.com/rds/>.
2. Clique em Launch a DB Instance (Iniciar uma instância de banco de dados).
3. Escolha Meu SQL e clique em Selecionar.
4. Deixe a seleção padrão de Implantação Multi-AZ e IOPS Armazenamento Provisionado e clique em Avançar.
5. Deixe as especificações da instância em seus valores padrão, especifique configurações e clique em Next (Próximo).
6. Na página Configure Advanced Settings, escolha um nome de security group apropriado e o nome do banco de dados. O grupo de segurança que você usa deve pelo menos permitir o TCP acesso de entrada para a porta 3306 a partir do nó principal do seu cluster. Se você não tiver criado o cluster neste ponto, poderá permitir que todos os hosts se conectem à porta 3306 e ajustar o security group depois de ter executado o cluster. Clique em Launch DB Instance (Iniciar instância de banco de dados).
7. No RDS Painel, selecione Instâncias e selecione a instância que você acabou de criar. Quando seu banco de dados estiver disponível, anote o nome do banco de dados, o nome de usuário, a senha e o nome do host da RDS instância. Você usará essas informações ao criar e configurar o cluster.


Para especificar um SQL banco de dados My externo para o Hue ao iniciar um cluster usando o AWS CLI

Para especificar um SQL banco de dados My externo para Hue ao iniciar um cluster usando o AWS CLI, use as informações que você observou ao criar sua RDS instância para configuração `hue.ini` com um objeto de configuração.

 Note

Você pode criar vários clusters que usam o mesmo banco de dados externo, mas cada cluster compartilhará informações de usuários e o histórico de usuários.

- Usando o AWS CLI, crie um cluster com o Hue instalado, usando o banco de dados externo que você criou e referenciando um arquivo de configuração com uma classificação de configuração para o Hue que especifica as propriedades do banco de dados. O exemplo a seguir cria um cluster com o Hue instalado, fazendo referência a um arquivo de configuração no Amazon S3, `myConfig.json`, que especifica a configuração do banco de dados.

 Note

Os caracteres de continuação de linha do Linux (`\`) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (`^`).

```
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
```

Os conteúdos de exemplo do arquivo `myConfig.json` são mostrados abaixo. Substituir *dbname*, *username*, *password* e *RDS instance hostname* com os valores que você anotou anteriormente no RDS Painel.

```
[{
  "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": []  
      },  
    ]  
  }  
]
```

Solução de problemas

Em caso de RDS falha da Amazon

É possível que os usuários encontrem atrasos ao executar uma consulta, pois a instância de banco de dados do Hue não está respondendo ou está em processo de failover. Veja a seguir estão alguns fatos e diretrizes para esse problema:

- Se você fizer login no RDS console da Amazon, poderá pesquisar eventos de failover. Por exemplo, para ver se um failover é em andamento ou já ocorreu, procure eventos como "Multi-AZ instance failover started" e "Multi-AZ instance failover completed".
- São necessários cerca de 30 segundos para que uma RDS instância conclua um failover.
- Se você estiver recebendo longer-than-normal respostas para consultas no Hue, tente reexecutar a consulta.

Configurações avançadas do Hue

Esta seção inclui os seguintes tópicos.

Tópicos

- [Configurar o Hue para usuários LDAP](#)

Configurar o Hue para usuários LDAP

A integração com LDAP permite que os usuários façam login no Hue usando as credenciais existentes armazenadas em um LDAP diretório. Ao integrar o Hue com LDAP, você não precisa gerenciar de forma independente as informações do usuário no Hue. As informações abaixo demonstram a integração do Hue com o Microsoft Active Directory, mas as opções de configuração são análogas a qualquer diretório. LDAP

LDAPa autenticação primeiro se liga ao servidor e estabelece a conexão. Em seguida, a conexão estabelecida é usada para qualquer consulta subsequente para pesquisar informações LDAP do usuário. A menos que o seu servidor do Active Directory permita conexões anônimas, uma conexão precisa ser estabelecida usando um nome diferenciado e uma senha de associação. O nome diferenciado (ou DN) de associação é definido pela configuração `bind_dn`. A senha de associação é definida pela configuração `bind_password`. O Hue tem duas maneiras de vincular LDAP solicitações: vinculação de pesquisa e associação direta. O método preferido para usar o Hue com a Amazon EMR é a vinculação de pesquisa.

Quando a associação de pesquisa é usada com o Active Directory, o Hue usa o atributo de nome de usuário (definido por `user_name_attr` config) para localizar o atributo que precisa ser recuperado do nome diferenciado (ou DN) base. A associação de pesquisa é útil quando o DN completo não é conhecido pelo usuário do Hue.

Por exemplo, você pode ter `user_name_attr` config definido para usar o nome comum (ou CN). Nesse caso, o servidor do Active Directory usa o nome de usuário do Hue fornecido durante o login para pesquisar a árvore de diretório em busca de um nome comum correspondente, começando com o nome diferenciado base. Se o nome comum do usuário do Hue for encontrado, o nome diferenciado desse usuário será retornado pelo servidor. Em seguida, o Hue constrói um nome diferenciado usado para autenticar o usuário, realizando uma operação de associação.

Note

A associação de pesquisa procura nomes de usuários em todas as subárvores de diretórios, começando com o nome diferenciado base. O nome distinto básico especificado na LDAP configuração do Hue deve ser o pai mais próximo do nome de usuário, ou seu desempenho de LDAP autenticação poderá ser prejudicado.

Quando a associação direta é usada com o Active Directory, o `nt_domain` ou `ldap_username_pattern` exato deve ser usado para se autenticar. Quando a associação direta é usada, se o atributo de domínio `nt` (definido pela configuração `nt_domain`) estiver especificado, um template de nome diferenciado do usuário será criado usando o formato: `<login username>@nt_domain`. Esse template é usado para pesquisar todas as subárvores de diretórios, começando com o nome diferenciado base. Se o domínio `nt` não for configurado, o Hue procurará um padrão de nome diferenciado exato para o usuário (definido pela configuração `ldap_username_pattern`). Nesse caso, o servidor procura um valor `ldap_username_pattern` correspondente em todas as subárvores de diretórios, começando com o nome diferenciado base.

Para iniciar um cluster com LDAP propriedades para o Hue usando o AWS CLI

- Para especificar LDAP propriedades para `hue-ini`, crie um cluster com o Hue instalado e faça referência a um arquivo json com propriedades de configuração para LDAP. Um exemplo de comando é mostrado abaixo, que faz referência a um arquivo de configuração `myConfig.json` armazenado no 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.
```

Os conteúdos de exemplo de `myConfig.json` são mostrados abaixo.

```
[
  {
    "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

Com a Amazon EMR versão 5.21.0 e posterior, você pode substituir as configurações do cluster e especificar classificações de configuração adicionais para cada grupo de instâncias em um cluster em execução. Você faz isso usando o EMR console da Amazon, o AWS Command Line Interface (AWS CLI) ou AWS SDK o. Para obter mais informações, consulte [Supplying a Configuration for an Instance Group in a Running Cluster](#).

Para visualizar LDAP as configurações no Hue

1. Verifique se você tem uma VPN conexão ativa ou um SSH túnel com o nó principal do EMR cluster da Amazon. Em seguida, no seu navegador, digite *master-public-dns*:8888 para abrir a interface web do Hue.
2. Faça login usando suas credenciais de administrador do Hue. Se a janela Did you know? (Você sabia?) for aberta, clique em Got it, prof! (Entendido!) para fechá-la.
3. Clique no ícone do Hue na barra de ferramentas.
4. Na página Sobre o Hue, selecione Configuração.
5. Na seção Configuration Sections and Variables (Seções e variáveis de configuração), clique em Desktop.
6. Role até a seção ldap para exibir as configurações.

Histórico de versões do Hue

A tabela a seguir lista a versão do Hue incluída em cada versão de lançamento da AmazonEMR, junto com os componentes instalados com o aplicativo. Para obter as versões dos componentes em cada versão, consulte a seção Versão de componentes da sua versão em [Versões de lançamento do Amazon EMR 7.x](#), [Versões de lançamento do Amazon EMR 6.x](#) ou [Versões de lançamento do Amazon EMR 5.x](#).

Informações de versão do Hue

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o Hue
emr-7.2.0	4.11.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, hadoop-cl

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o Hue
		ient, 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, mariadb-server, oozie-client, oozie-server

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o Hue
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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hue-server, mariadb-server, oozie-client, oozie-server

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o Hue
emr-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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hue-server, mysql-server, oozie-client, oozie-server

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Hue	Componentes instalados com o 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

O [Apache Iceberg](#) é um formato de tabela aberta para grandes conjuntos de dados no Amazon Simple Storage Service (Amazon S3). Ele fornece desempenho rápido de consultas em tabelas grandes, confirmações atômicas, gravações simultâneas e evolução de tabelas SQL compatíveis. A partir do Amazon EMR 6.5.0, você pode usar o Apache Spark 3 em EMR clusters da Amazon com o formato de tabela Iceberg.

A tabela a seguir lista a versão do Iceberg incluída na versão mais recente da série Amazon EMR 7.x, junto com os componentes que a Amazon EMR instala com o Iceberg.

Para a versão dos componentes instalados com o Iceberg nesta versão, consulte [Versões de componentes da versão 7.2.0](#).

Informações sobre a versão do Iceberg para o emr-7.2.0

Gravadora EMR de lançamento da Amazon	Versão do Iceberg	Componentes instalados com o Iceberg
emr-7.2.0	Iceberg 1.5.0-amzn-0	Not available.

A tabela a seguir lista a versão do Iceberg incluída na versão mais recente da série EMR 6.x da Amazon, junto com os componentes que a Amazon EMR instala com o Iceberg.

Para obter a versão dos componentes instalados com o Iceberg nessa versão, consulte [Release 6.15.0 Component Versions](#).

Informações de versão do Iceberg para o emr-6.15.0

Gravadora EMR de lançamento da Amazon	Versão do Iceberg	Componentes instalados com o Iceberg
emr-6.15.0	Iceberg 1.4.0-amzn-0	Not available.

Tópicos

- [Como o Iceberg funciona](#)
- [Usar um cluster com o Iceberg instalado](#)

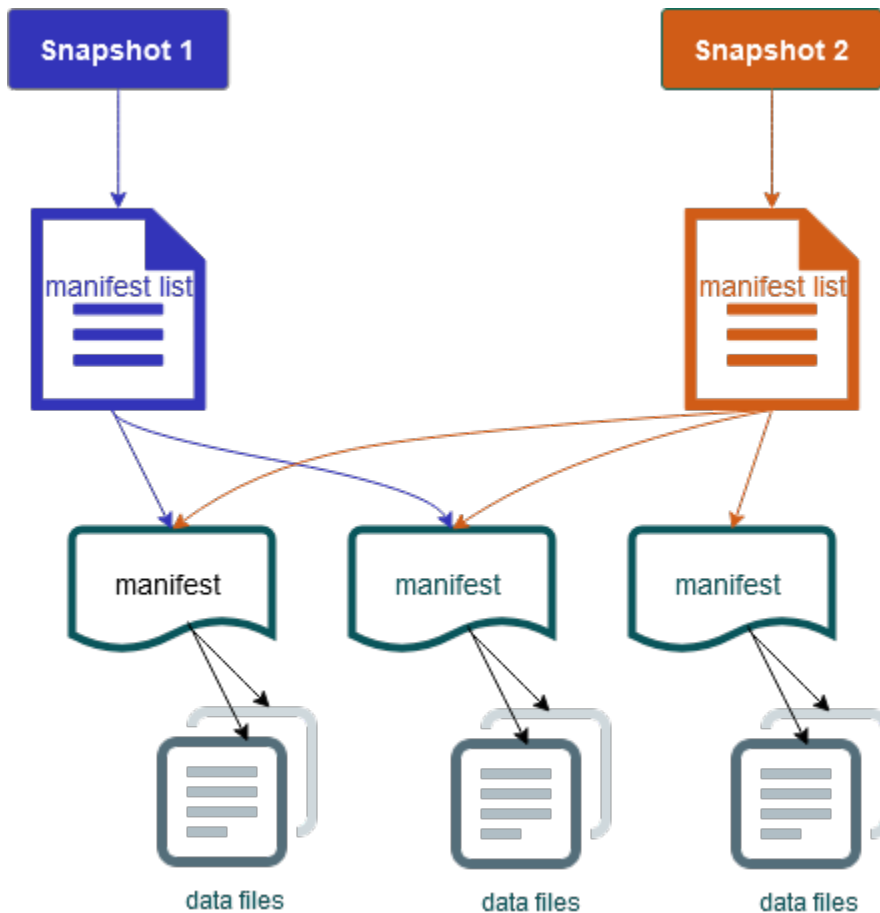
- [Considerações e limitações para usar o Iceberg na Amazon EMR](#)
- [Histórico de versões do Iceberg](#)

Como o Iceberg funciona

O Iceberg rastreia arquivos de dados individuais em uma tabela em vez de em diretórios. Dessa forma, os gravadores podem criar arquivos de dados no local (os arquivos não são movidos nem alterados). Além disso, os gravadores só podem adicionar arquivos à tabela em uma confirmação explícita. O estado da tabela é mantido em arquivos de metadados. Todas as alterações no estado da tabela criam um novo arquivo de metadados que substitui atômicamente os metadados antigos. O arquivo de metadados da tabela rastreia o esquema da tabela, a configuração do particionamento e outras propriedades.

Ele também inclui snapshots do conteúdo da tabela. Cada snapshot é um conjunto completo de arquivos de dados na tabela em um momento específico. Os snapshots são listados no arquivo de metadados, mas os arquivos de um snapshot são armazenados em arquivos manifesto separados. As transições atômicas de um arquivo de metadados da tabela para o próximo fornecem isolamento de snapshots. Os leitores usam o instantâneo que estava atualizado quando carregaram os metadados da tabela. Os leitores não são afetados pelas alterações até atualizarem e escolherem um novo local de metadados. Arquivos de dados em snapshots são armazenados em um ou mais arquivos manifesto que contêm uma linha para cada arquivo de dados na tabela, seus dados de partição e suas métricas. Um snapshot é a união de todos os arquivos nos seus manifestos. Os arquivos manifesto também podem ser compartilhados entre snapshots para evitar a regravação de metadados que são alterados com pouca frequência.

Diagrama de snapshots do Iceberg



O Iceberg oferece os seguintes atributos:

- Oferece suporte a ACID transações e viagens no tempo em seu data lake do Amazon S3.
- As novas tentativas de confirmação se beneficiam das vantagens de performance da [simultaneidade otimista](#).
- A resolução de conflitos em nível de arquivo resulta em alta simultaneidade.
- Com estatísticas mínimas e máximas por coluna nos metadados, você pode pular arquivos, o que aumenta a performance de consultas seletivas.
- Você pode organizar tabelas em layouts de partição flexíveis, com a evolução da partição permitindo atualizações nos esquemas de partição. As consultas e os volumes de dados podem, então, mudar sem contar com diretórios físicos.
- Oferece suporte à [evolução e à aplicação do esquema](#).
- As tabelas do Iceberg funcionam como coletores idempotentes e fontes reproduzíveis. Isso permite streaming e suporte em lote com pipelines exatamente uma vez. Os coletores idempotentes rastreiam operações de gravação que foram bem-sucedidas no passado. Portanto, o coletor pode

solicitar dados novamente em caso de falha e descartar dados que tiverem sido enviados várias vezes.

- Visualize o histórico e a linhagem, incluindo evolução da tabela, o histórico de operações e as estatísticas de cada confirmação.
- Migre de um conjunto de dados existente com uma opção de formato de dados (ParquetORC, Avro) e mecanismo de análise (Spark, Trino, PrestoDB, Flink, Hive).

Usar um cluster com o Iceberg instalado

Esta seção inclui informações sobre como usar o Iceberg com Spark, Trino, Flink e Hive.

Usar um cluster do Iceberg com o Spark

A partir da EMR versão 6.5.0 da Amazon, você pode usar o Iceberg com seu cluster Spark sem a necessidade de incluir ações de bootstrap. Para EMR as versões 6.4.0 e anteriores da Amazon, você pode usar uma ação de bootstrap para pré-instalar todas as dependências necessárias.

Neste tutorial, você usa o AWS CLI para trabalhar com o Iceberg em um cluster do Amazon EMR Spark. Para usar o console para criar um cluster com o Iceberg instalado, siga as etapas em [Criar um data lake do Apache Iceberg usando Amazon Athena, Amazon](#) e Glue. EMR AWS

Criar um cluster do Iceberg

Você pode criar um cluster com o Iceberg instalado usando o AWS Management Console, o AWS CLI ou o Amazon EMR API. Neste tutorial, você usa o AWS CLI para trabalhar com o Iceberg em um EMR cluster da Amazon. Para usar o console para criar um cluster com o Iceberg instalado, siga as etapas em [Criar um data lake do Apache Iceberg usando Amazon Athena, Amazon](#) e Glue. EMR AWS

Para usar o Iceberg na Amazon EMR com o AWS CLI, primeiro crie um cluster com as etapas a seguir. Para obter informações sobre como especificar a classificação do Iceberg usando o AWS CLI, consulte [Forneça uma configuração usando o AWS CLI ao criar um cluster](#) ou [Forneça uma configuração usando o Java SDK ao criar um cluster](#)

1. Crie um arquivo `configurations.json`, com o seguinte conteúdo:

```
[{
  "Classification": "iceberg-defaults",
```

```
"Properties":{"iceberg.enabled":"true"}
}]
```

2. Em seguida, crie um cluster com a configuração a seguir. Substitua o exemplo do caminho do bucket do Amazon S3 e o ID da sub-rede pelos seus.

```
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
```

Como alternativa, você pode criar um EMR cluster da Amazon incluindo o aplicativo Spark e incluir o arquivo `/usr/share/aws/iceberg/lib/iceberg-spark3-runtime.jar` como uma JAR dependência em um trabalho do Spark. Para obter mais informações, consulte [Submitting Applications](#).

Para incluir o jar como uma dependência em um trabalho do Spark, adicione a seguinte propriedade de configuração à aplicação do Spark:

```
--conf "spark.jars=/usr/share/aws/iceberg/lib/iceberg-spark3-runtime.jar"
```

Para obter mais informações sobre as dependências de trabalho do Spark, consulte [Gerenciamento de dependências](#) no documento do Apache Spark [Executar o Spark no Kubernetes](#).

Inicializar uma sessão do Spark para Iceberg

Os exemplos a seguir demonstram como iniciar o shell interativo do Spark, usar o envio do Spark ou usar o Amazon EMR Notebooks para trabalhar com o Iceberg na Amazon. EMR

spark-shell

1. Conecte-se ao nó principal usando SSH o. Para obter mais informações, consulte [Conecte-se ao nó principal usando SSH](#) o Amazon EMR Management Guide.

2. Digite o seguinte comando para iniciar o shell do Spark. Para usar a PySpark concha, `spark-shell` substitua por `spark-shell`.

```
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. Conecte-se ao nó principal usando SSH o. Para obter mais informações, consulte [Conecte-se ao nó principal usando SSH](#) o Amazon EMR Management Guide.
2. Insira o comando a seguir para iniciar a sessão do Spark no 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

Para inicializar uma sessão do Spark usando notebooks EMR Studio, configure sua sessão do Spark usando o comando `%%configure` mágico em seu EMR notebook Amazon, como no exemplo a seguir. Para obter mais informações, consulte [Use EMR Notebooks magics](#) no Amazon EMR Management Guide.

```
%%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/"
}
}
```

Gravar em uma tabela do Iceberg

O exemplo a seguir mostra como criar um DataFrame e gravá-lo como um conjunto de dados do Iceberg. Os exemplos demonstram como trabalhar com conjuntos de dados usando o shell do Spark enquanto estão conectados ao nó principal, usando SSH como usuário padrão do hadoop.

Note

Para colar exemplos de código no shell do Spark, digite `:paste` no prompt, cole o exemplo e pressione CTRL+D.

PySpark

O Spark inclui um shell baseado em Python, `pyspark`, que você pode usar para gerar protótipos de programas Spark escritos em Python. Invoque `pyspark` no nó 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()
```

Ler em uma tabela do 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
```

Configure as propriedades do Spark para usar o AWS Glue Data Catalog como metastore de tabelas Iceberg

Para usar o AWS Glue Catalog como Metastore para tabelas Iceberg, defina as propriedades de configuração do Spark conforme abaixo:

```
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
```

Usar um cluster do Iceberg com o Trino

A partir da EMR versão 6.6.0 da Amazon, você pode usar o Iceberg com seu cluster Trino.

Neste tutorial, você usa o AWS CLI para trabalhar com o Iceberg em um cluster Amazon EMR Trino. Para usar o console para criar um cluster com o Iceberg instalado, siga as etapas em [Criar um data lake do Apache Iceberg usando Amazon Athena, Amazon](#) e Glue. EMR AWS

Criar um cluster do Iceberg

Para usar o Iceberg na Amazon EMR com o AWS CLI, primeiro crie um cluster com as etapas a seguir. Para obter informações sobre como especificar a classificação do Iceberg usando o AWS CLI, consulte [Forneça uma configuração usando o AWS CLI ao criar um cluster](#) ou [Forneça uma configuração usando o Java SDK ao criar um cluster](#)

1. Crie um arquivo `iceberg.properties` e defina um valor para o catálogo escolhido. Por exemplo, se você quiser usar a metastore do Hive como catálogo, o arquivo deverá ter o conteúdo a seguir.

```
connector.name=iceberg  
hive.metastore.uri=thrift://localhost:9083
```

Se você quiser usar o AWS Glue Data Catalog como sua loja, seu arquivo deve ter o seguinte conteúdo.

```
connector.name=iceberg  
iceberg.catalog.type=glue
```

2. Crie uma ação de bootstrap que copie `iceberg.properties` do Amazon S3 para `/etc/trino/conf/catalog/iceberg.properties`, como no exemplo a seguir. Para obter mais

informações sobre ações de bootstrap, consulte [Criar ações de bootstrap para instalar softwares adicionais](#).

```
set -ex
sudo aws s3 cp s3://DOC-EXAMPLE-BUCKET/iceberg.properties /etc/trino/conf/catalog/iceberg.properties
```

3. Crie um cluster com a configuração a seguir, substituindo o exemplo de caminho de script de ações de bootstrap e o nome da chave pelos seus.

```
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>
```

Inicializar uma sessão do Trino para Iceberg

Para inicializar uma sessão no Trino, execute o comando a seguir.

```
trino-cli --catalog iceberg
```

Gravar em uma tabela do Iceberg

Crie e grave na sua tabela com os SQL comandos a seguir.

```
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');
```

Ler em uma tabela do Iceberg

Para ler a tabela do Iceberg, execute o comando a seguir.

```
trino> SELECT * from default.iceberg_table;
```

Usar um cluster do Iceberg com o Flink

A partir da EMR versão 6.9.0 da Amazon, você pode usar o Iceberg com um cluster Flink sem as etapas de configuração necessárias ao usar a integração Iceberg Flink de código aberto.

Criar um cluster no Iceberg

Você pode criar um cluster com o Iceberg instalado usando o AWS Management Console AWS CLI, o ou o Amazon EMR API. Neste tutorial, você usa o AWS CLI para trabalhar com o Iceberg em um EMR cluster da Amazon. Para usar o console para criar um cluster com o Iceberg instalado, siga as etapas em [Criar um data lake do Apache Iceberg usando Amazon Athena, Amazon](#) e Glue. EMR AWS

Para usar o Iceberg na Amazon EMR com o AWS CLI, primeiro crie um cluster com as etapas a seguir. Para obter informações sobre como especificar a classificação do Iceberg usando o AWS CLI, consulte [Forneça uma configuração usando o AWS CLI ao criar um cluster](#) ou [Forneça uma configuração usando o Java SDK ao criar um cluster](#) Crie um arquivo denominado `configurations.json` com o seguinte conteúdo:

```
[{
  "Classification":"iceberg-defaults",
  "Properties":{"iceberg.enabled":"true"}
}]
```

Em seguida, crie um cluster com a configuração a seguir, substituindo o exemplo de caminho do bucket do Amazon S3 e o ID da sub-rede pelos seus próprios valores:

```
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
```

Como alternativa, você pode criar um cluster Amazon EMR 6.9.0 com um aplicativo Flink e usar o arquivo `/usr/share/aws/iceberg/lib/iceberg-flink-runtime.jar` como uma JAR dependência em um trabalho do Flink.

Usando o cliente Flink SQL

O script SQL do cliente está localizado em `/usr/lib/flink/bin`. Você pode executar o script com o seguinte comando:

```
flink-yarn-session -d # starting the Flink YARN Session in detached mode  
./sql-client.sh
```

Isso lança um Flink SQL Shell.

Exemplos do Flink

Criar uma tabela no 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);
```

Tabela 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);");
```

Gravar em uma tabela do Iceberg

Flink SQL

```
INSERT INTO `glue_catalog`.`<DB>`.`sample` values (1, 'a'),(2,'b'),(3,'c');
```

Tabela API

```
tEnv.executeSql(
    "INSERT INTO `glue_catalog`.`"
```

```
+ db
+ ``.`sample` values (1, 'a'),(2,'b'),(3,'c');");
```

Fluxo de dados 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 glueCatalogLoader =
    CatalogLoader.custom(
        "glue",
        props,
        new Configuration(),
        "org.apache.iceberg.aws.glue.GlueCatalog");

TableLoader tableLoader =
    TableLoader.fromCatalog(glueCatalogLoader, TableIdentifier.of(db, "sample"));

DataStreamSink<Void> dataStreamSink =
    FlinkSink.forRowData(input).tableLoader(tableLoader).append();

env.execute("Datastream Write");
```

Ler em uma tabela do Iceberg

Flink SQL

```
SELECT * FROM `glue_catalog`.`<DB>`.`sample`;
```

Tabela API

```
Table result = tEnv.sqlQuery("select * from `glue_catalog`.`" + db + "`.`sample`");
```

Fluxo de dados 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");
```

Usar o catálogo do Hive

Certifique-se de que as dependências do Flink e do Hive sejam resolvidas conforme descrito em [Configurar o Flink com o Hive Metastore e o Catálogo do Glue](#).

Executar um trabalho do Flink

Uma forma de enviar um trabalho para o Flink é usar uma sessão do Flink YARN por trabalho. Isso pode ser iniciado com o seguinte comando:

```
sudo flink run -m yarn-cluster -p 4 -yjm 1024m -ytm 4096m $JAR_FILE_NAME
```

Usar um cluster do Iceberg com o Hive

Com as EMR versões 6.9.0 e superiores da Amazon, você pode usar o Iceberg com um cluster Hive sem precisar executar as etapas de configuração necessárias para a integração de código aberto com o Iceberg Hive. Para EMR as versões 6.8.0 e anteriores da Amazon, você pode usar uma ação de bootstrap para instalar o `iceberg-hive-runtime jar` e configurar o suporte ao Hive for Iceberg.

O Amazon EMR 6.9.0 inclui todos os recursos para a [integração do Hive 3.1.3 com o Iceberg 0.14.1](#) e também inclui recursos EMR adicionais da Amazon, como seleção automática de mecanismos de execução compatíveis em tempo de execução (Amazon na versão 6.9.0). EMR EKS

Criar um cluster do Iceberg

Você pode criar um cluster com o Iceberg instalado usando o AWS Management Console, o AWS CLI ou o Amazon EMR API. Neste tutorial, você usa o AWS CLI para trabalhar com o Iceberg em um EMR cluster da Amazon. Para usar o console para criar um cluster com o Iceberg instalado, siga as etapas em [Criar um data lake do Iceberg usando Amazon Athena, EMR AWS Amazon](#) e Glue.

Para usar o Iceberg na Amazon EMR com o AWS CLI, primeiro crie um cluster usando as etapas abaixo. Para obter informações sobre como especificar a classificação Iceberg usando o AWS CLI ou o Java SDK, consulte [Forneça uma configuração usando o AWS CLI ao criar um cluster](#) ou [Forneça uma configuração usando o Java SDK ao criar um cluster](#). Crie um arquivo chamado `configurations.json` com o seguinte conteúdo:

```
[{
  "Classification":"iceberg-defaults",
  "Properties":{"iceberg.enabled":"true"}
}]
```

Em seguida, crie um cluster com a configuração a seguir, substituindo o exemplo de caminho do bucket do Amazon S3 e o ID da sub-rede pelos seus próprios:

```
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
```

Um cluster do Iceberg no Hive faz o seguinte:

- Carrega o jar de runtime do Hive no Iceberg no Hive e habilita a configuração relacionada ao Iceberg para o mecanismo do Hive.
- Permite a seleção dinâmica do mecanismo de execução do Amazon EMR Hive para impedir que os usuários configurem um mecanismo de execução compatível com o Iceberg.

Note

Atualmente, os clusters Hive Iceberg não oferecem suporte ao AWS Glue Data Catalog. O catálogo padrão do Iceberg é HiveCatalog, que corresponde ao metastore configurado para o ambiente Hive. Para obter mais informações sobre gerenciamento de catálogos, consulte [Usando HCatalog](#) na documentação do [Apache Hive](#).

Suporte a recursos

O Amazon EMR 6.9.0 é compatível com o Hive 3.1.3 e o Iceberg 0.14.1. O suporte a atributos é limitado aos atributos compatíveis com o Iceberg para o Hive 3.1.2 e 3.1.3. Os seguintes comandos são compatíveis:

- Com as EMR versões 6.9.0 a 6.12.x da Amazon, você deve incluir o libfb303 jar no diretório do Hive. `auxlib` Use o seguinte comando para instalar:

```
sudo /usr/bin/ln -sf /usr/lib/hive/lib/libfb303-*.jar /usr/lib/hive/auxlib/  
libfb303.jar
```

Com as EMR versões 6.13 e posteriores da Amazon, o libfb303 jar é automaticamente vinculado ao diretório Hiveauxlib.


- Criar uma tabela

- Tabela não particionada: as tabelas externas no Hive podem ser criadas ao ser fornecido o manipulador de armazenamento da seguinte forma:

```
CREATE EXTERNAL TABLE x (i int) STORED BY  
'org.apache.iceberg.mr.hive.HiveIcebergStorageHandler'
```

- Tabela particionada: as tabelas externas particionadas no Hive podem ser criadas da seguinte forma:

```
CREATE EXTERNAL TABLE x (i int) PARTITIONED BY (j int) STORED BY  
'org.apache.iceberg.mr.hive.HiveIcebergStorageHandler'
```

 Note

O formato de STORED AS arquivo de ORC/AVRO/não PARQUET é suportado no Hive 3. A opção padrão e única é Parquet.

- Descartar uma tabela: o comando DROP TABLE é usado para descartar tabelas, como no exemplo a seguir:

```
DROP TABLE [IF EXISTS] table_name [PURGE];
```

- Ler uma tabela: as instruções SELECT podem ser usadas para ler tabelas do Iceberg no Hive, como no exemplo a seguir. Os mecanismos de execução compatíveis são MR e Tez.

```
SELECT * FROM table_name
```

[Para obter informações sobre a sintaxe de seleção do Hive, consulte *LanguageManual Selecionar*.](#)

Para obter informações sobre instruções selecionadas com tabelas do Iceberg no Hive, consulte [Selecionar Apache Iceberg](#).

- Inserir em uma tabela: a instrução INSERT INTO do HiveQL funciona em tabelas do Iceberg compatíveis somente com o mecanismo de execução do Map Reduce. EMROs usuários da Amazon não precisam definir explicitamente o mecanismo de execução porque o Amazon EMR Hive seleciona o mecanismo para tabelas Iceberg em tempo de execução.

- Inserção de tabela única: exemplo:

```
INSERT INTO table_name VALUES ('a', 1);  
INSERT INTO table_name SELECT...;
```

- Inserção de várias tabelas: a inserção de várias tabelas não atômicas em instruções é compatível. Exemplo:

```
FROM source  
INSERT INTO table_1 SELECT a, b  
INSERT INTO table_2 SELECT c,d;
```

Considerações e limitações para usar o Iceberg na Amazon EMR

Esta seção inclui considerações e limitações sobre como usar o Iceberg com Spark, Trino, Flink e Hive.

Considerações sobre o uso do Iceberg com o Spark

- O Amazon EMR 6.5.0 não oferece suporte ao Iceberg executado EMR na Amazon EKS por padrão. Uma imagem personalizada do Amazon EMR 6.5.0 está disponível para que você possa passar `--jars local:///usr/share/aws/iceberg/lib/iceberg-spark3-runtime.jar` como `spark-submit` parâmetro para criar tabelas Iceberg na Amazon EMR em EKS. Para obter mais informações, consulte [Enviar uma carga de trabalho do Spark na Amazon EMR usando uma imagem personalizada](#) no Amazon EMR on EKS Development Guide. Você também pode entrar em contato com AWS Support para obter assistência. A partir do Amazon EMR 6.6.0, o Iceberg é suportado na Amazon on. EMR EKS
- Ao usar o AWS Glue como um catálogo para o Iceberg, verifique se o banco de dados no qual você está criando uma tabela existe no AWS Glue. Se você estiver usando serviços como AWS Lake Formation e não conseguir carregar o catálogo, verifique se você tem acesso adequado ao serviço para executar o comando.

Considerações sobre o uso do Iceberg com o Trino

- O Amazon EMR 6.5 não oferece suporte nativo do catálogo Trino Iceberg para Iceberg. O Trino precisa do Iceberg v0.11, então recomendamos lançar um EMR cluster Amazon para Trino separado do cluster Spark e incluir o Iceberg v0.11 nesse cluster.

- Ao usar o AWS Glue como um catálogo para o Iceberg, certifique-se de que o banco de dados no qual você está criando uma tabela exista no AWS Glue. Se você estiver usando serviços como AWS Lake Formation e não conseguir carregar o catálogo, verifique se você tem acesso adequado ao serviço para executar o comando.

Considerações sobre o uso do Iceberg com o Flink

Ao usar o AWS Glue como um catálogo para o Iceberg, verifique se o banco de dados no qual você está criando uma tabela existe no AWS Glue. Se você estiver usando serviços como AWS Lake Formation e não conseguir carregar o catálogo, verifique se você tem acesso adequado ao serviço para executar o comando.

Considerações sobre o uso do Iceberg com o Hive

- O Iceberg é compatível com os seguintes tipos de consulta:
 - Create table
 - Drop table
 - Inserir na tabela
 - Ler a tabela
- Somente o mecanismo de execução MR (MapReduce) é suportado para operações DML (linguagem de manipulação de dados), e o MR está obsoleto no Hive 3.1.3.
- AWS Atualmente, o Glue Data Catalog não é compatível com o Iceberg with Hive.
- O tratamento de erros não é suficientemente robusto. Em casos de configuração incorreta, as inserções nas consultas podem ser concluídas com êxito. No entanto, a falha na atualização dos metadados pode resultar em perda de dados.

Histórico de versões do Iceberg

A tabela a seguir lista a versão do Iceberg incluída em cada versão de lançamento da AmazonEMR, junto com os componentes instalados com o aplicativo. Para obter as versões dos componentes em cada versão, consulte a seção Versão de componentes da sua versão em [Versões de lançamento](#)

[do Amazon EMR 7.x](#), [Versões de lançamento do Amazon EMR 6.x](#) ou [Versões de lançamento do Amazon EMR 5.x](#).

Informações sobre a versão do Iceberg

Gravadora Amazon EMR Release	Versão do Iceberg	Componentes instalados com o Iceberg
emr-7.2.0	1.5.0-aman-0	Not available.
emr-7.1.0	1.4.3-amzn-0	Not available.
emr-7.0.0	q1.4.2-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.
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.

Gravadora Amazon EMR Release	Versão do Iceberg	Componentes instalados com o Iceberg
emr-6.5.0	0.12.0	Not available.

Notas da versão do Iceberg por versão

- [Amazon EMR 6.9.0 - Notas de lançamento do Iceberg](#)

Amazon EMR 6.9.0 - Notas de lançamento do Iceberg

Amazon EMR 6.9.0 - Mudanças no Iceberg

Tipo	Descrição
Atributo	Integração EMR do Amazon Flink com o Iceberg.
Atributo	Integração do Amazon EMR Hive com o Iceberg.
Atributo	Support para armazenar arquivos de metadados do Iceberg no Amazon for Lustre FSx para melhorar o tempo de planejamento da consulta.
Backport	PR 5050 : Flink 1.15: Support opções de gravação nos comentários de inserção em linha. SQL
Backport	PR 5282 AWS: Corrija falhas de PUT repetição abrindo novos fluxos de arquivos de dados.
Backport	PR 5318 : Flink 1.15: Preenche a lacuna entre FlinkSource e IcebergSource (FLIP-27) e adicionou uma configuração opcional para usar a fonte -27 no Flink. FLIP SQL

Tipo	Descrição
Backport	PR 5344 : Flink 1.14: Preenche a lacuna entre FlinkSource e IcebergSource (FLIP-27) e adicionou uma configuração opcional para usar a fonte -27 no Flink. FLIP SQL
Backport	PR 5393 : Flink 1.14, 1.15: Evite converter métricas Iceberg MetricContext em Flink no leitor de origem -27. FLIP
Backport	PR 5401 : Flink 1.14, 1.15: IcebergSourceReader Grupo perdido no PR #5393 para FLIP -27 métricas do leitor de origem.
Backport	PR 5679 : Spark 3.2, 3.3: Corrige a propagação de nulidade para o nó. MergeRows
Backport	PR 5860 : Spark 3.3: Correção QueryFailure ao executar RewriteManifestProcedure em tabelas particionadas por data.
Backport	PR 5880 : Spark 3.3: Corrige a nulidade nas projeções. merge-on-read
Backport	PR 5917 : Spark 3.2: Corrige a nulidade nas projeções. merge-on-read

Caderno Jupyter no Amazon EMR

O [Caderno Jupyter](#) é uma aplicação Web de código aberto que você pode usar para criar e compartilhar documentos que contêm código ativo, equações, visualizações e texto narrativo. O Amazon EMR oferece três opções para trabalhar com cadernos Jupyter:

Tópicos

- [EMR Studio](#)
- [Caderno do Amazon EMR baseado no Caderno Jupyter](#)
- [JupyterHub](#)

EMR Studio

O Amazon EMR Studio é um ambiente de desenvolvimento integrado (IDE) baseado na Web para [cadernos Jupyter](#) totalmente gerenciados que são executados em clusters do Amazon EMR. Você pode configurar um EMR Studio para sua equipe desenvolver, visualizar e depurar aplicações escritas em R, Python, Scala e PySpark.

Recomendamos usar o EMR Studio ao usar notebooks Jupyter no Amazon EMR. Para obter mais informações, consulte [EMR Studio](#) no Guia de gerenciamento do Amazon EMR.

Caderno do Amazon EMR baseado no Caderno Jupyter

Os Cadernos do EMR constituem um ambiente do [Caderno Jupyter](#) incorporado ao console do Amazon EMR que permite criar rapidamente cadernos Jupyter, anexá-los a clusters do Spark e, em seguida, abrir o editor do Caderno Jupyter no console para executar consultas e códigos remotamente. Um caderno do EMR é salvo no Amazon S3 independentemente dos clusters para armazenamento durável, acesso rápido e flexibilidade. Você pode ter vários cadernos abertos, conectar vários cadernos a um único cluster e reutilizar um caderno em clusters diferentes.

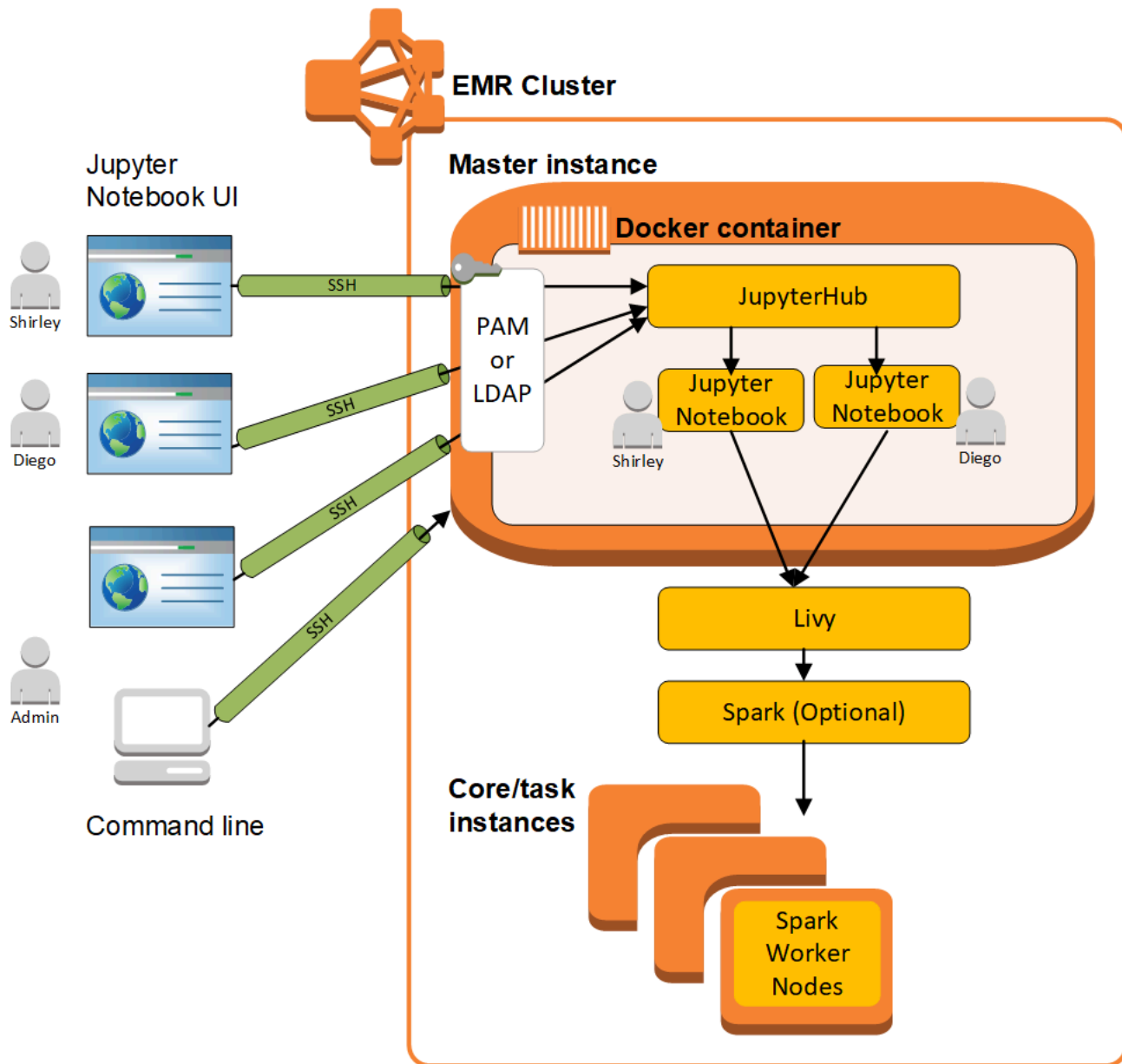
Para obter mais informações, consulte [Cadernos do EMR](#) no Guia de gerenciamento do Amazon EMR.

JupyterHub

O [Caderno Jupyter](#) é uma aplicação Web de código aberto que você pode usar para criar e compartilhar documentos que contêm código ativo, equações, visualizações e texto narrativo. [JupyterHub](#) permite que você hospede várias instâncias de um servidor de notebook Jupyter de usuário único. Quando você cria um cluster com JupyterHub, a Amazon EMR cria um contêiner Docker no nó principal do cluster. JupyterHub, todos os componentes necessários para o Jupyter e o [Sparkmagic](#) funcionam dentro do contêiner.

O Sparkmagic é uma biblioteca de kernels que permite que os notebooks Jupyter interajam com o Apache [Spark executado na Amazon EMR por meio de, que é um servidor para o Spark](#). [Apache Livy](#) REST O Spark e o Apache Livy são instalados automaticamente quando você cria um cluster com o. JupyterHub O kernel padrão do Python 3 para o Jupyter está disponível junto com os kernels PySpark 3, PySpark, e Spark que estão disponíveis com o Sparkmagic. Você pode usar esses kernels para executar código ad-hoc do Spark e consultas SQL interativas usando Python e Scala. Você pode instalar kernels adicionais dentro do contêiner Docker manualmente. Para ter mais informações, consulte [Instalar Kernels e bibliotecas adicionais](#).

O diagrama a seguir mostra os componentes do JupyterHub na Amazon EMR com os métodos de autenticação correspondentes para usuários de notebooks e administradores. Para ter mais informações, consulte [Adicionar usuários e administradores do Caderno Jupyter](#).



A tabela a seguir lista a versão JupyterHub incluída na versão mais recente da série Amazon EMR 7.x, junto com os componentes com os quais a Amazon é EMR instalada. JupyterHub

Para a versão dos componentes instalados JupyterHub nesta versão, consulte [Versões de componentes da versão 7.2.0](#).

JupyterHub informações sobre a versão do emr-7.2.0

Gravadora EMR de lançamento da Amazon	JupyterHub Versão	Componentes instalados com 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

A tabela a seguir lista a versão JupyterHub incluída na versão mais recente da série EMR 6.x da Amazon, junto com os componentes com os quais a Amazon é EMR instalada. JupyterHub

Para a versão dos componentes instalados JupyterHub nesta versão, consulte [Versões de componentes da versão 6.15.0](#).

JupyterHub informações da versão do emr-6.15.0

Gravadora EMR de lançamento da Amazon	JupyterHub Versão	Componentes instalados com 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

Gravadora EMR de lançamento da Amazon	JupyterHub Versão	Componentes instalados com JupyterHub
		, hudi, hudi-spark, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, livy-server, jupyterhub

A tabela a seguir lista a versão JupyterHub incluída na versão mais recente da série Amazon EMR 5.x, junto com os componentes com os quais a Amazon é EMR instalada. JupyterHub

Para a versão dos componentes instalados JupyterHub nesta versão, consulte Versões de componentes da [versão 5.36.2](#).

JupyterHub informações da versão do emr-5.36.2

Gravadora EMR de lançamento da Amazon	JupyterHub Versão	Componentes instalados com 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

O kernel do Python 3 incluído na JupyterHub Amazon EMR é o 3.6.4.

As bibliotecas instaladas no `jupyterhub` contêiner podem variar entre as versões de EMR lançamento da Amazon e as EC2 AMI versões da Amazon.

Para listar bibliotecas instaladas usando o **conda**

- Execute o seguinte comando na linha de comandos do nó principal:

```
sudo docker exec jupyterhub bash -c "conda list"
```

Para listar bibliotecas instaladas usando o **pip**

- Execute o seguinte comando na linha de comandos do nó principal:

```
sudo docker exec jupyterhub bash -c "pip freeze"
```

Tópicos

- [Crie um cluster com JupyterHub](#)
- [Considerações ao usar JupyterHub na Amazon EMR](#)
- [Configurando JupyterHub](#)
- [Configurar a persistência de cadernos no Amazon S3](#)
- [Conectar-se ao nó principal e aos servidores de cadernos](#)
- [JupyterHub configuração e administração](#)
- [Adicionar usuários e administradores do Caderno Jupyter](#)
- [Instalar Kernels e bibliotecas adicionais](#)
- [JupyterHub histórico de lançamentos](#)

Crie um cluster com JupyterHub

Você pode criar um EMR cluster da Amazon JupyterHub usando o AWS Management Console AWS Command Line Interface, ou o Amazon EMR API. Certifique-se de que o cluster não seja criado com a opção para encerrar automaticamente após concluir etapas (opção `--auto-terminate` na AWS CLI). Além disso, certifique-se de que os administradores de notebook e os usuários possam acessar o par de chaves que você usa ao criar o cluster. Para obter mais informações, consulte [Use a key pair for SSH credenciais](#) no Amazon EMR Management Guide.

Crie um cluster JupyterHub usando o console

Use o procedimento a seguir para criar um cluster JupyterHub instalado usando Opções avançadas no EMR console da Amazon.

Para criar um EMR cluster da Amazon com JupyterHub instalado usando o EMR console da Amazon

1. Navegue até o novo EMR console da Amazon e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte [Usar o console antigo](#).
2. Escolha Create cluster (Criar cluster), Go to advanced options (Ir para opções avançadas).
3. Em Software Configuration (Configuração de software):
 - Em Release, selecione emr-5.36.2 e escolha. JupyterHub
 - Se você usa o Spark, para usar o AWS Glue Data Catalog como metastore do SparkSQL, selecione Usar para metadados da tabela do Spark. Para ter mais informações, consulte [Use o AWS Glue Data Catalog como metastore para o Spark SQL](#).
 - Em Editar configurações de software, escolha Inserir configuração e especificar valores ou escolha Carregar JSON do S3 e especifique um arquivo JSON de configuração. Para ter mais informações, consulte [Configurando JupyterHub](#).
4. Em (Add steps (optional)) Adicionar etapas (opcional) configure as etapas para serem executadas quando o cluster for criado, certifique-se de que Auto-terminate cluster after the last step is completed (Encerrar o cluster automaticamente após a última etapa for concluída) não esteja selecionada e escolha Next (Próximo).
5. Escolha a opção Hardware Configuration (Configuração de hardware), Next (Próximo). Para obter mais informações, consulte [Configurar o hardware e a rede do cluster](#) no Amazon EMR Management Guide.
6. Escolha opções para General Cluster Settings (Configurações gerais do cluster), Next (Próximo).
7. Escolha Security Options (Opções de segurança), especificando um par de chaves e escolha Create Cluster (Criar cluster).

Crie um cluster JupyterHub usando o AWS CLI

Para iniciar um cluster com JupyterHub, use o `aws emr create-cluster` comando e, para a `--applications` opção, especifique `Name=JupyterHub`. O exemplo a seguir inicia um JupyterHub cluster na Amazon EMR com duas EC2 instâncias (uma instância principal e uma instância principal).

Além disso, a depuração é habilitada, com os logs armazenados no local do Amazon S3 conforme especificado pelo `--log-uri`. O par de chaves especificado fornece acesso às EC2 instâncias da Amazon no cluster.

Note

Os caracteres de continuação de linha do Linux (`\`) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (`^`).

```
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
```

Considerações ao usar JupyterHub na Amazon EMR

Considere o seguinte ao usar JupyterHub na AmazonEMR.

- #### Warning

Os notebooks e arquivos do usuário são salvos no sistema de arquivos no nó principal. Este é um armazenamento temporário que não persiste por meio da finalização do cluster. Quando um cluster é finalizado, esses dados são perdidos se não forem feitos backup. Recomendamos que você programe backups regulares usando trabalhos `cron` ou outro meio adequado para seu aplicativo.

Além disso, as alterações de configuração feitas no contêiner podem não persistir se o contêiner for reiniciado. Recomendamos que você faça script ou automatize a configuração de contêiner para que possa reproduzir personalizações com mais facilidade.
- A autenticação Kerberos que foi configurada usando uma configuração de EMR segurança da Amazon não é suportada.
- Não há suporte ao [OAuthenticator](#).

Configurando JupyterHub

Você pode personalizar a configuração de notebooks JupyterHub na Amazon EMR e de usuários individuais conectando-se ao nó principal do cluster e editando os arquivos de configuração. Depois de alterar os valores, reinicie o contêiner `jupyterhub`.

Modifique as propriedades nos arquivos a seguir para configurar JupyterHub notebooks Jupyter individuais:

- `jupyterhub_config.py`: por padrão, esse arquivo é salvo no diretório `/etc/jupyter/conf/` no nó principal. Para obter mais informações, consulte [Princípios básicos de configuração](#) na JupyterHub documentação.
- `jupyter_notebook_config.py`: esse arquivo é salvo no diretório `/etc/jupyter/` por padrão e copiado para o contêiner `jupyterhub` como padrão. Para obter mais informações, consulte [Arquivo de configuração e opções de linha de comando](#) na documentação do Notebook Jupyter.

Você também pode usar a classificação de configuração `jupyter-sparkmagic-conf` para personalizar o Sparkmagic, que atualiza valores no arquivo `config.json` para o Sparkmagic. Para obter mais informações sobre as configurações disponíveis, consulte [example_config.json](#) em GitHub. Para obter mais informações sobre o uso de classificações de configuração com aplicativos na AmazonEMR, consulte [Configurar aplicações](#).

O exemplo a seguir inicia um cluster usando o AWS CLI, referenciando o arquivo `MyJupyterConfig.json` para as configurações de classificação do Sparkmagic.

Note

Os caracteres de continuação de linha do Linux (`\`) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (`^`).

```
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
```

Os conteúdos de exemplo de `MyJupyterConfig.json` são os seguintes:

```
[
  {
    "Classification": "jupyter-sparkmagic-conf",
    "Properties": {
      "kernel_python_credentials" : "{\"username\":\"diego\", \"base64_password\": \"mypass\", \"url\":\"http://localhost:8998\", \"auth\":\"None\"}"
    }
  }
]
```

Note

Com a Amazon EMR versão 5.21.0 e posterior, você pode substituir as configurações do cluster e especificar classificações de configuração adicionais para cada grupo de instâncias em um cluster em execução. Você faz isso usando o EMR console da Amazon, o AWS Command Line Interface (AWS CLI) ou AWS SDK o. Para obter mais informações, consulte [Supplying a Configuration for an Instance Group in a Running Cluster](#).

Configurar a persistência de cadernos no Amazon S3

Você pode configurar um JupyterHub cluster na Amazon EMR para que os notebooks salvos por um usuário permaneçam no Amazon S3, fora do armazenamento temporário em instâncias de cluster. EC2

Especifique a persistência do Amazon S3 usando a classificação de configuração `jupyter-s3-conf` ao criar um cluster. Para ter mais informações, consulte [Configurar aplicações](#).

Além de habilitar a persistência do Amazon S3 usando a propriedade `s3.persistence.enabled`, especifique um bucket no Amazon S3 no qual os cadernos são salvos com a propriedade `s3.persistence.bucket`. Os notebooks para cada usuário são salvos em uma pasta `jupyter/jupyterhub-user-name` no bucket especificado. O bucket já deve existir no Amazon S3, e a função do perfil de EC2 instância que você especifica ao criar o cluster deve ter permissões para o bucket (por padrão, a função é `EMR_EC2_DefaultRole`). Para obter mais informações, consulte [Configurar IAM funções para EMR permissões de AWS serviços da Amazon](#).

Quando você inicia um novo cluster usando a mesma classificação de configuração propriedades, os usuários podem abrir notebooks com o conteúdo do local salvo.

Observe que, ao importar arquivos como módulos em um caderno quando o Amazon S3 está habilitado, isso resulta no carregamento dos arquivos para o Amazon S3. Quando você importa arquivos sem habilitar a persistência do Amazon S3, eles são enviados para o seu contêiner.

JupyterHub

O exemplo a seguir habilita a persistência do Amazon S3. Os notebooks salvos por usuários são salvos na pasta `s3://MyJupyterBackups/jupyter/jupyterhub-user-name` para cada usuário, em que *jupyterhub-user-name* é um nome de usuário, como `diego`.

```
[
  {
    "Classification": "jupyter-s3-conf",
    "Properties": {
      "s3.persistence.enabled": "true",
      "s3.persistence.bucket": "MyJupyterBackups"
    }
  }
]
```

Conectar-se ao nó principal e aos servidores de cadernos

JupyterHub administradores e usuários de notebooks devem se conectar ao nó principal do cluster usando um SSH túnel e, em seguida, conectar-se às interfaces da web servidas pelo JupyterHub nó principal. Para obter mais informações sobre como configurar um SSH túnel e usar o túnel para conexões proxy da Web, consulte [Connect to the cluster](#) no Amazon EMR Management Guide.

Por padrão, JupyterHub na Amazon EMR está disponível por meio da porta 9443 no nó principal. O JupyterHub proxy interno também atende às instâncias do notebook por meio da porta 9443. JupyterHub e as interfaces web do Jupyter podem ser acessadas usando um URL com o seguinte padrão:

`https://MasterNodeDNS: 9443`

Você pode especificar uma porta diferente usando a propriedade `c.JupyterHub.port` no arquivo `jupyterhub_config.py`. Para obter mais informações, consulte [Noções básicas sobre redes](#) na JupyterHub documentação.

Por padrão, JupyterHub na Amazon EMR usa um certificado autoassinado para uso HTTPS de SSL criptografia. Os usuários são solicitados a confiar no certificado autoassinado quando eles se conectarem. Você pode usar um certificado confiável e suas próprias chaves.

Substitua o arquivo de certificado padrão, `server.crt`, e arquivo chave `server.key` no diretório `/etc/jupyter/conf/` no nó principal com o seu próprio certificado e arquivos de chave. Use as `c.JupyterHub.ssl_cert` propriedades `c.JupyterHub.ssl_key` e no `jupyterhub_config.py` arquivo para especificar seus SSL materiais. Para obter mais informações, consulte [Configurações de segurança](#) na JupyterHub documentação. Depois de atualizar o `jupyterhub_config.py`, reinicie o contêiner.

JupyterHub configuração e administração

JupyterHub e componentes relacionados são executados dentro de um contêiner Docker chamado `jupyterhub` que executa o sistema operacional Ubuntu. Há várias maneiras para você administrar os componentes que são executados dentro do contêiner.

Warning

As personalizações que você executa no contêiner podem não persistir se o contêiner for reiniciado. Recomendamos que você faça script ou automatize a configuração de contêiner para que possa reproduzir personalizações com mais facilidade.

Administração usando a linha de comando

Quando conectado ao nó principal usando SSH, você pode emitir comandos usando a interface de linha de comando do Docker (CLI) e especificando o contêiner por nome (`jupyterhub`) ou ID. Por exemplo, o `sudo docker exec jupyterhub command` executa comandos reconhecidos pelo sistema operacional ou por um aplicativo em execução dentro do contêiner. Você pode usar esse método para adicionar usuários ao sistema operacional e instalar aplicativos e bibliotecas adicionais no contêiner Docker. Por exemplo, a imagem do contêiner padrão inclui Conda para instalação de pacote, portanto, você pode executar o seguinte comando na linha de comando do nó principal para instalar um aplicativo, Keras, dentro do contêiner:

```
sudo docker exec jupyterhub conda install keras
```

Administração por envio de etapas

As etapas são uma maneira de enviar trabalhos a um cluster. Você pode enviar etapas quando o cluster é iniciado ou pode enviá-las para um cluster em execução. Os comandos que você executa

na linha de comando podem ser enviados como etapas usando o `command-runner.jar`. Para obter mais informações, consulte [Trabalhe com etapas usando o console CLI](#) e no Guia EMR de Gerenciamento da Amazon [Executar comandos e scripts em um cluster do Amazon EMR](#) e.

Por exemplo, você pode usar o AWS CLI comando a seguir em um computador local para instalar o Keras da mesma forma que fez na linha de comando do nó principal no exemplo anterior:

```
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"
```

Além disso, você pode fazer o script de uma sequência de etapas, carregar o script no Amazon S3 e, em seguida usar `script-runner.jar` para executar o script quando você cria o cluster ou adiciona o script como uma etapa. Para ter mais informações, consulte [Executar comandos e scripts em um cluster do Amazon EMR](#). Para ver um exemplo, consulte [the section called “Exemplo: script Bash para adicionar vários usuários”](#).

Administração usando REST APIs

Jupyter, JupyterHub, e o HTTP proxy para JupyterHub fornecer REST APIs que você pode usar para enviar solicitações. Para enviar solicitações para JupyterHub, você deve passar um API token com a solicitação. Você pode usar o `curl` comando da linha de comando do nó principal para executar REST comandos. Para obter mais informações, consulte os seguintes recursos do :

- O [uso JupyterHub está REST API](#) na documentação de JupyterHub, que inclui instruções para gerar API tokens
- [Servidor Jupyter Notebook ativado API](#) GitHub
- [configurable-http-proxy](#) em GitHub

O exemplo a seguir demonstra o uso do REST API for JupyterHub para obter uma lista de usuários. O comando passa um token administrativo gerado anteriormente e usa a porta padrão, 9443 JupyterHub, para canalizar a saída para `jq` para facilitar a visualização:

```
curl -XGET -s -k https://$HOST:9443/hub/api/users \
-H "Authorization: token $admin_token" | jq .
```

Adicionar usuários e administradores do Caderno Jupyter

Você pode usar um dos dois métodos para os usuários se autenticarem para que JupyterHub possam criar cadernos e, opcionalmente, administrar. JupyterHub O método mais fácil é usar o módulo JupyterHub de autenticação conectável (PAM). Além disso, a JupyterHub Amazon EMR oferece suporte ao [plug-in LDAP autenticador JupyterHub para](#) obter identidades de usuário de um LDAP servidor, como um servidor Microsoft Active Directory. Instruções e exemplos para adicionar usuários com cada método de autenticação são fornecidos nesta seção.

JupyterHub na Amazon EMR tem um usuário padrão com permissões de administrador. O nome de usuário é `jovyan` e a senha é `jupyter`. É altamente recomendável que você substitua o usuário por outro usuário que tenha permissões administrativas. Você pode fazer isso usando uma etapa ao criar o cluster ou se conectando ao nó principal quando o cluster estiver em execução.

Tópicos

- [Utilizar a autenticação PAM](#)
- [Utilizar a autenticação LDAP](#)
- [Personificação do usuário](#)

Utilizar a autenticação PAM

Criar PAM usuários JupyterHub na Amazon EMR é um processo de duas etapas. A primeira etapa é adicionar usuários ao sistema operacional em execução no contêiner `jupyterhub` no nó principal e adicionar um diretório inicial do usuário correspondente para cada usuário. A segunda etapa é adicionar esses usuários do sistema operacional como JupyterHub usuários — um processo conhecido como lista branca em. JupyterHub Depois que um JupyterHub usuário é adicionado, ele pode se conectar ao JupyterHub URL e fornecer suas credenciais do sistema operacional para acesso.

Quando um usuário faz login, JupyterHub abre a instância do servidor do notebook para esse usuário, que é salva no diretório inicial do usuário no nó principal, que é `/var/lib/jupyter/home/username`. Se uma instância do servidor do notebook não existir, JupyterHub gera uma instância do notebook no diretório inicial do usuário. As seções a seguir demonstram como adicionar usuários individualmente ao sistema operacional e ao sistema operacional JupyterHub, seguido por um script bash rudimentar que adiciona vários usuários.

Adicionar um usuário do sistema operacional ao contêiner

O exemplo a seguir usa primeiro o comando [useradd](#) dentro do contêiner para adicionar um único usuário, diego e criar um diretório inicial para esse usuário. O segundo comando usa [chpasswd](#) para estabelecer uma senha de diego para este usuário. Os comandos são executados na linha de comando do nó principal enquanto estão conectados usando SSH. Você também pode executar esses comandos usando uma etapa conforme descrito anteriormente em [Administração por envio de etapas](#).

```
sudo docker exec jupyterhub useradd -m -s /bin/bash -N diego
sudo docker exec jupyterhub bash -c "echo diego:diego | chpasswd"
```

Adicionando um JupyterHub usuário

Você pode usar o painel de administração em JupyterHub ou REST API para adicionar usuários e administradores, ou apenas usuários.

Para adicionar usuários e administradores usando o painel de administração no JupyterHub

1. Conecte-se ao nó principal usando SSH e faça login em <https://MasterNodeDNS:9443> com uma identidade que tenha permissões de administrador.
2. Escolha Control Panel (Painel de Controle), Admin.
3. Escolha User (Usuário), Add Users (Adicionar usuários) ou escolha Admin, Add Admins (Adicionar admins).

Para adicionar um usuário usando o REST API

1. Conecte-se ao nó principal usando SSH e use o comando a seguir no nó principal ou execute o comando como uma etapa.
2. Adquira um token administrativo para fazer API solicitações e substituir *AdminToken* na etapa seguinte com esse token.
3. Use o comando a seguir, substituindo *UserName* com um usuário do sistema operacional que foi criado dentro do contêiner.

```
curl -XPOST -H "Authorization: token AdminToken" "https://$(hostname):9443/hub/api/users/UserName"
```

Note

Você é automaticamente adicionado como usuário JupyterHub não administrador ao fazer login na interface da JupyterHub web pela primeira vez.

Exemplo: script Bash para adicionar vários usuários

O exemplo de script bash a seguir reúne as etapas anteriores desta seção para criar vários JupyterHub usuários. O script pode ser executado diretamente no nó principal, ou pode ser carregado no Amazon S3 e, em seguida, executado como uma etapa.

O script primeiro estabelece uma matriz de nomes de usuário e usa o `jupyterhub token` comando para criar um API token para o administrador padrão, `jovyan`. Em seguida, ele cria um usuário do sistema operacional no contêiner `jupyterhub` para cada usuário, atribuindo uma senha inicial a cada um que é igual ao seu nome de usuário. Por fim, ele chama a REST API operação para criar cada usuário em JupyterHub. Ele passa o token gerado anteriormente no script e canaliza a REST resposta `jq` para facilitar a visualização.

```
# 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
```

Salve o script em um local do Amazon S3, como `s3://mybucket/createjupyterusers.sh`. Em seguida, você pode usar o `script-runner.jar` para executá-lo como uma etapa.

Exemplo: executar o script ao criar um cluster (AWS CLI)

Note

Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

```
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"]
```

Executar o script em um cluster existente (AWS CLI)

Note

Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

```
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"]
```

Utilizar a autenticação LDAP

O Lightweight Directory Access Protocol (LDAP) é um protocolo de aplicativo para consultar e modificar objetos que correspondem a recursos, como usuários e computadores, armazenados em um provedor de serviços LDAP de diretório compatível, como o Active Directory ou um servidor aberto. LDAP Você pode usar o [plug-in LDAP autenticador](#) do JupyterHub with JupyterHub on Amazon EMR LDAP para usar na autenticação do usuário. O plug-in gerencia as sessões de login LDAP dos usuários e fornece informações do usuário ao Jupyter. Isso permite que os usuários

se conectem a notebooks JupyterHub e notebooks usando as credenciais de suas identidades armazenadas em um LDAP servidor compatível.

As etapas desta seção orientam você pelas etapas a seguir para configurar e ativar o LDAP uso do plug-in LDAP autenticador para JupyterHub. Execute as etapas enquanto estiver conectado à linha de comando do nó principal. Para ter mais informações, consulte [Conectar-se ao nó principal e aos servidores de cadernos](#).

1. Crie um arquivo de LDAP configuração com informações sobre o LDAP servidor, como endereço IP do host, porta, nomes de associação e assim por diante.
2. Modifique `/etc/jupyter/conf/jupyterhub_config.py` para habilitar o plug-in LDAP autenticador para JupyterHub.
3. Crie e execute um script que seja configurado LDAP dentro do `jupyterhub` contêiner.
4. Consulte LDAP os usuários e, em seguida, crie diretórios pessoais dentro do contêiner para cada usuário. JupyterHub requer diretórios pessoais para hospedar notebooks.
5. Execute um script que reinicia JupyterHub

Important

Antes de configurar LDAP, teste sua infraestrutura de rede para garantir que o LDAP servidor e o nó principal do cluster possam se comunicar conforme necessário. TLS normalmente usa a porta 389 em uma TCP conexão simples. Se sua LDAP conexão usa SSL, a TCP porta conhecida SSL é 636.

Crie o arquivo LDAP de configuração

O exemplo abaixo usa os seguintes valores de configuração do espaço reservado. Substitua esses valores pelos parâmetros que corresponde à sua implementação.

- O LDAP servidor está executando a versão 3 e está disponível na porta 389. Esta é a não SSL porta padrão para LDAP.
- O nome diferenciado base (DN) é `dc=example, dc=org`.

Use um editor de texto para criar o arquivo [ldap.conf](#), com conteúdo semelhante ao seguinte. Use valores apropriados para sua LDAP implementação. Substituir *host* com o endereço IP ou nome de host resolvível do seu LDAP servidor.

```
base dc=example,dc=org
uri ldap://host
ldap_version 3
binddn cn=admin,dc=example,dc=org
bindpw admin
```

Ativar o plug-in LDAP autenticador para JupyterHub

Use um editor de texto para modificar o arquivo `/etc/jupyter/conf/jupyterhub_config.py` e adicionar as propriedades [ldapauthenticator](#) semelhante ao seguinte. Substituir *host* com o endereço IP ou nome de host resolvível do LDAP servidor. O exemplo pressupõe que os objetos do usuário estejam dentro de uma unidade organizacional (ou) chamada *people*, e usa os componentes de nomes distintos que você estabeleceu anteriormente usando `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'
```

Configurar LDAP dentro do contêiner

Use um editor de texto para criar um script bash com o seguinte conteúdo:

```
#!/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"
```

Salve o script no nó principal e, em seguida, execute-o na linha de comando do nó principal. Por exemplo, com o script salvo como `configure_ldap_client.sh`, torne o arquivo executável:

```
chmod +x configure_ldap_client.sh
```

E execute o script:

```
./configure_ldap_client.sh
```

Adicionar atributos ao Active Directory

Para encontrar cada usuário e criar a entrada apropriada no banco de dados, o contêiner JupyterHub docker requer as seguintes UNIX propriedades para o objeto de usuário correspondente no Active Directory. Para obter mais informações, consulte a seção Como continuo editando os atributosGID/UIDRFC2307 agora que o Plug-in de Atributos Unix não está mais disponível para o snap-in Usuários e Computadores do Active Directory? MMC no artigo [Esclarecimento sobre o status do gerenciamento de identidade para Unix \(IDMU\) e função de NIS servidor na versão prévia técnica do Windows Server 2016 e versões posteriores.](#)

- `homeDirectory`

Este é o local para o diretório inicial do usuário, que é geralmente `/home/username`.

- `gidNumber`

Um valor maior que 60000 que ainda não esteja sendo usado por outro usuário. Verifique o arquivo `etc/passwd` para ver os GIDs que estão sendo usados.

- `uidNumber`

Um valor maior que 60000 que ainda não esteja sendo usado por outro grupo. Verifique o arquivo `etc/group` para ver os UDIs que estão sendo usados.

- `uid`

Isso é o mesmo que o `username`.

Criar diretórios iniciais do usuário

JupyterHub precisa de diretórios iniciais dentro do contêiner para autenticar LDAP usuários e armazenar dados da instância. O exemplo a seguir demonstra dois usuários, shirley e diego, no diretório. LDAP

A primeira etapa é consultar o LDAP servidor para obter as informações de ID de usuário e ID de grupo de cada usuário usando [ldapsearch](#), conforme mostrado no exemplo a seguir, substituindo `host` com o endereço IP ou nome de host resolvível do seu LDAP servidor:

```
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
```

O `ldapsearch` comando retorna uma resposta em LDIF formato -que se parece com a seguinte para os usuários shirley e 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
```

Ao usar as informações da resposta, execute comandos dentro do contêiner para criar um diretório inicial para cada nome comum de usuário (cn). Use o `uidNumber` e o `gidNumber` para corrigir a propriedade do diretório inicial desse usuário. Os comandos de exemplo a seguir fazem isso para o usuário *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

LDAPO autenticador para JupyterHub não oferece suporte à criação de usuários locais. Para obter mais informações, consulte a [nota de configuração do LDAP autenticador sobre a criação de usuários locais](#).

Para criar um usuário local manualmente, use o comando a seguir.

```
sudo docker exec jupyterhub bash -c "echo 'shirley:x:$uidNumber:$gidNumber::/home/shirley:/bin/bash' >> /etc/passwd"
```

Reinicie o JupyterHub contêiner

Para reiniciar o contêiner `jupyterhub`, execute os seguintes comandos:

```
sudo docker stop jupyterhub
sudo docker start jupyterhub
```

Personificação do usuário

Um trabalho do Spark executado dentro de um notebook Jupyter percorre vários aplicativos durante sua execução na Amazon. EMR Por exemplo, PySpark 3 códigos que um usuário executa dentro do Jupyter são recebidos pelo Sparkmagic, que usa uma HTTP POST solicitação para enviá-los ao Livy, que então cria um trabalho do Spark para executar no cluster usando YARN.

Por padrão, os YARN trabalhos enviados dessa forma são executados como usuário `livy`, independentemente do usuário que iniciou o trabalho. Ao configurar a representação do usuário, você pode fazer com que o ID do usuário do notebook também seja o usuário associado ao YARN trabalho. Em vez dos trabalhos serem iniciados por `shirley` e `diego` associados ao usuário `livy`, os trabalhos que cada usuário inicia são associados a `shirley` e `diego` respectivamente. Isso ajuda você a auditar o uso do Jupyter e gerenciar os aplicativos em sua organização.

Essa configuração tem suporte somente quando as chamadas do Sparkmagic para o Livy são não autenticadas. Os aplicativos que fornecem uma camada de proxy ou de autenticação entre os aplicativos Hadoop e Livy (como o Apache Knox Gateway) não são compatíveis. As etapas para configurar a representação do usuário nesta seção pressupõem que JupyterHub e Livy estejam sendo executados no mesmo nó principal. Se seu aplicativo tiver clusters separados, ele [Etapa 3: criar HDFS diretórios pessoais para usuários](#) precisará ser modificado para que os HDFS diretórios sejam criados no nó principal do Livy.

Etapas para configurar a personificação do usuário

- [Etapa 1: configurar o Livy](#)
- [Etapa 2: adicionar usuários](#)
- [Etapa 3: criar HDFS diretórios pessoais para usuários](#)

Etapa 1: configurar o Livy

Use as classificações de configuração `livy-conf` e `core-site` ao criar um cluster para habilitar a personificação do usuário do Livy, conforme mostrado no exemplo a seguir. Salve a classificação de

configuração como uma JSON e, em seguida, faça referência a ela ao criar o cluster ou especifique a classificação de configuração em linha. Para ter mais informações, consulte [Configurar aplicações](#).

```
[
  {
    "Classification": "livy-conf",
    "Properties": {
      "livy.impersonation.enabled": "true"
    }
  },
  {
    "Classification": "core-site",
    "Properties": {
      "hadoop.proxyuser.livy.groups": "*",
      "hadoop.proxyuser.livy.hosts": "*"
    }
  }
]
```

Etapa 2: adicionar usuários

Adicione JupyterHub usuários usando PAM ou LDAP. Para obter mais informações, consulte [Utilizar a autenticação PAM](#) e [Utilizar a autenticação LDAP](#).

Etapa 3: criar HDFS diretórios pessoais para usuários

Você se conectou ao nó principal para criar usuários. Ainda conectado a ele, copie o conteúdo abaixo e o salve em um arquivo de script. O script cria diretórios HDFS iniciais para cada JupyterHub usuário no nó principal. O script pressupõe que você esteja usando o ID de usuário administrador padrão, *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
```

Instalar Kernels e bibliotecas adicionais

Quando você cria um cluster JupyterHub na AmazonEMR, o kernel Python 3 padrão para Jupyter junto com os kernels Spark PySpark e Spark para Sparkmagic são instalados no contêiner Docker. Você pode instalar kernels adicionais. Você também pode instalar bibliotecas e pacotes adicionais e importá-los para o shell apropriado.

Instalar um Kernel

Os kernels são instalados no contêiner Docker. A maneira mais fácil de fazer isso é criar um script bash com comandos de instalação, salvá-lo no nó principal e, em seguida, usar o comando `sudo docker exec jupyterhub script_name` para executar o script dentro do contêiner `jupyterhub`. O script de exemplo a seguir instala o kernel e, em seguida, instala algumas bibliotecas para esse kernel no nó principal, para que mais tarde você possa importar as bibliotecas usando o kernel no 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
```

Para instalar o kernel e as bibliotecas dentro do contêiner, abra uma conexão de terminal com o nó principal, salve o script em `/etc/jupyter/install_kernels.sh` e execute o seguinte comando na linha de comando do nó principal:

```
sudo docker exec jupyterhub bash /etc/jupyter/install_kernels.sh
```

Usar bibliotecas e instalar bibliotecas adicionais

Um conjunto básico de bibliotecas de aprendizado de máquina e ciência de dados para Python 3 está pré-instalado na JupyterHub Amazon. EMR Você pode usar o `sudo docker exec jupyterhub bash -c "conda list"` e o `sudo docker exec jupyterhub bash -c "pip freeze"`.

Se um job Spark precisar de bibliotecas nos nós de operador, recomendamos que você use uma ação de bootstrap para executar um script para instalar as bibliotecas ao criar o cluster. As ações de bootstrap são executadas em todos os nós do cluster durante o processo de criação do cluster, o que simplifica a instalação. Se você instalar bibliotecas em nós core/de operador depois que um cluster estiver em execução, a operação será mais complexa. Nós fornecemos um exemplo de programa em Python nesta seção que mostra como instalar essas bibliotecas.

A ação de bootstrap e os exemplos de programas em Python mostrados nesta seção usam um script bash salvo no Amazon S3 para instalar as bibliotecas em todos os nós.

O script referenciado nos exemplos a seguir usa `pip` para instalar `paramiko`, `nltk`, `scipy`, `scikit-learn` e `pandas` para o kernel do Python 3:


```
#!/bin/bash

sudo python3 -m pip install boto3 paramiko nltk scipy scikit-learn pandas
```

Depois de criar o script, carregue-o em um local no Amazon S3, por exemplo, `s3://mybucket/install-my-jupyter-libraries.sh`. Para obter mais informações, consulte [Carregar objetos](#) no Guia do usuário do Amazon Simple Storage Service para que você possa usá-lo na sua ação de bootstrap ou no seu programa do Python.

Para especificar uma ação de bootstrap que instala bibliotecas em todos os nós ao criar um cluster usando o AWS CLI

1. Crie um script semelhante ao do exemplo anterior e salve-o em um local no Amazon S3. Nós usamos o exemplo `s3://mybucket/install-my-jupyter-libraries.sh`.
2. Crie o cluster com JupyterHub e use o Path argumento da `--bootstrap-actions` opção para especificar a localização do script, conforme mostrado no exemplo a seguir:

 Note

Os caracteres de continuação de linha do Linux (`\`) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (`^`).

```
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
```

Como especificar uma ação de bootstrap que instala bibliotecas em todos os nós ao criar um cluster usando o console

1. Navegue até o novo EMR console da Amazon e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte [Usar o console antigo](#).
2. Escolha Create cluster (Criar cluster), Go to advanced options (Ir para opções avançadas).
3. Especifique as configurações para Software and Steps (Software e etapas) e Hardware conforme apropriado para seu aplicativo.

4. Na tela General Cluster Settings (Configurações gerais de cluster), expanda Bootstrap Actions (Ações de bootstrap).
5. Para Add bootstrap action (Adicionar ação de bootstrap), selecione Custom action (Ação personalizada), Configure and add (Configurar e adicionar).
6. Em Nome, insira um nome fácil de lembrar. Em Localização do script, insira a localização do seu script no Amazon S3 (o exemplo que usamos é `s3://mybucket/install-my-jupyter-libraries.sh`). Deixe Optional arguments (Argumentos opcionais) em branco e escolha Add (Adicionar).
7. Especifique outras configurações para o seu cluster e escolha Next (Próximo).
8. Especifique as configurações de segurança e escolha Create cluster (Criar cluster).

Example Instalar bibliotecas em nós centrais de um cluster em execução

Após instalar bibliotecas no nó principal de dentro do Jupyter, você pode instalar bibliotecas em nós core em execução de várias maneiras. O exemplo a seguir mostra um programa Python escrito para ser executado em uma máquina local. Quando você executa o programa Python localmente, ele usa o `AWS-RunShellScript` of AWS Systems Manager para executar o script de exemplo, mostrado anteriormente nesta seção, que instala bibliotecas nos nós principais do 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 histórico de lançamentos

A tabela a seguir lista a versão JupyterHub incluída em cada versão de lançamento da AmazonEMR, junto com os componentes instalados com o aplicativo. Para obter as versões dos componentes em cada versão, consulte a seção Versão de componentes da sua versão em [Versões de lançamento do Amazon EMR 7.x](#), [Versões de lançamento do Amazon EMR 6.x](#) ou [Versões de lançamento do Amazon EMR 5.x](#).

JupyterHub informações sobre a versão

Gravadora Amazon EMR Release	JupyterHub Versão	Componentes instalados com 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-

Gravadora Amazon EMR Release	JupyterHub Versão	Componentes instalados com JupyterHub
		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-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

Gravadora Amazon EMR Release	JupyterHub Versão	Componentes instalados com 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
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

Gravadora Amazon EMR Release	JupyterHub Versão	Componentes instalados com 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
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

Gravadora Amazon EMR Release	JupyterHub Versão	Componentes instalados com 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
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

Gravadora Amazon EMR Release	JupyterHub Versão	Componentes instalados com 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
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

Gravadora Amazon EMR Release	JupyterHub Versão	Componentes instalados com 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
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

Gravadora Amazon EMR Release	JupyterHub Versão	Componentes instalados com 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
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

Gravadora Amazon EMR Release	JupyterHub Versão	Componentes instalados com 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
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

Gravadora Amazon EMR Release	JupyterHub Versão	Componentes instalados com 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
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

Gravadora Amazon EMR Release	JupyterHub Versão	Componentes instalados com 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
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

Gravadora Amazon EMR Release	JupyterHub Versão	Componentes instalados com 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
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

Gravadora Amazon EMR Release	JupyterHub Versão	Componentes instalados com 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
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

Gravadora Amazon EMR Release	JupyterHub Versão	Componentes instalados com 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
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

Gravadora Amazon EMR Release	JupyterHub Versão	Componentes instalados com 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
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

Gravadora Amazon EMR Release	JupyterHub Versão	Componentes instalados com 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
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

Gravadora Amazon EMR Release	JupyterHub Versão	Componentes instalados com 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
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

Gravadora Amazon EMR Release	JupyterHub Versão	Componentes instalados com 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
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

Gravadora Amazon EMR Release	JupyterHub Versão	Componentes instalados com 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
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

Gravadora Amazon EMR Release	JupyterHub Versão	Componentes instalados com 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
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

Gravadora Amazon EMR Release	JupyterHub Versão	Componentes instalados com 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
emr-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

Gravadora Amazon EMR Release	JupyterHub Versão	Componentes instalados com 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
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

Gravadora Amazon EMR Release	JupyterHub Versão	Componentes instalados com 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
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

Gravadora Amazon EMR Release	JupyterHub Versão	Componentes instalados com 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
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

Gravadora Amazon EMR Release	JupyterHub Versão	Componentes instalados com 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
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

Gravadora Amazon EMR Release	JupyterHub Versão	Componentes instalados com 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
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

Gravadora Amazon EMR Release	JupyterHub Versão	Componentes instalados com 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
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

Gravadora Amazon EMR Release	JupyterHub Versão	Componentes instalados com 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
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

Gravadora Amazon EMR Release	JupyterHub Versão	Componentes instalados com 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
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

Gravadora Amazon EMR Release	JupyterHub Versão	Componentes instalados com 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
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

Gravadora Amazon EMR Release	JupyterHub Versão	Componentes instalados com 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
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

Gravadora Amazon EMR Release	JupyterHub Versão	Componentes instalados com 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
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

Gravadora Amazon EMR Release	JupyterHub Versão	Componentes instalados com 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
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

Gravadora Amazon EMR Release	JupyterHub Versão	Componentes instalados com 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
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

Gravadora Amazon EMR Release	JupyterHub Versão	Componentes instalados com 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
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

Gravadora Amazon EMR Release	JupyterHub Versão	Componentes instalados com 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
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

O Livy permite a interação em uma REST interface com um EMR cluster executando o Spark. Você pode usar a REST interface ou uma biblioteca RPC cliente para enviar trabalhos do Spark ou trechos de código do Spark, recuperar resultados de forma síncrona ou assíncrona e gerenciar o Spark Context. Para obter mais informações, consulte o [site do Apache Livy](#). Livy está incluído na EMR versão 5.9.0 e posterior da Amazon.

Para acessar a interface web do Livy, configure um SSH túnel para o nó principal e uma conexão proxy. Para obter mais informações, consulte [Exibir interfaces web hospedadas em EMR clusters](#).

A tabela a seguir lista a versão do Livy incluída na versão mais recente da série Amazon EMR 7.x, junto com os componentes que a Amazon EMR instala com o Livy.

Para a versão dos componentes instalados com o Livy nesta versão, consulte [Versões de componentes da versão 7.2.0](#).

Informações sobre a versão do Livy para o emr-7.2.0

Gravadora EMR de lançamento da Amazon	Versão do Livy	Componentes instalados com o 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

A tabela a seguir lista a versão do Livy incluída na versão mais recente da série Amazon EMR 6.x, junto com os componentes que a Amazon EMR instala com o Livy.

Para obter a versão dos componentes instalados com o Livy nessa versão, consulte [Release 6.15.0 Component Versions](#).

Informações de versão do Livy para o emr-6.15.0

Gravadora EMR de lançamento da Amazon	Versão do Livy	Componentes instalados com o 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

A tabela a seguir lista a versão do Livy incluída na versão mais recente da série Amazon EMR 5.x, junto com os componentes que a Amazon EMR instala com o Livy.

Para a versão dos componentes instalados com o Livy nesta versão, consulte Versões de componentes da [versão 5.36.2](#).

Informações sobre a versão do Livy para o emr-5.36.2

Gravadora EMR de lançamento da Amazon	Versão do Livy	Componentes instalados com o 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-

Gravadora EMR de lançamento da Amazon	Versão do Livy	Componentes instalados com o 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

Tópicos

- [Habilitando HTTPS com o Apache Livy](#)
- [Histórico de versões do Livy](#)

Habilitando HTTPS com o Apache Livy

1. Provisione um EMR cluster da Amazon com criptografia de trânsito ativada. Para saber mais sobre criptografia, consulte [Criptografar dados em repouso e em trânsito](#).
2. Crie um arquivo chamado `livy_ssl.sh` com o conteúdo a seguir.

```
#!/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. Execute o script a seguir como uma EMR etapa da Amazon. Esse script é modificado `/etc/livy/conf/livy.conf` para ser ativado SSL.

```
--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. Reinicialize o serviço Apache Livy para que as alterações surtem efeito. Para reiniciar o Apache Livy, consulte [Interromper e reiniciar processos](#).
5. Teste se os clientes agora podem se comunicar usando HTTPS. Para enviar um trabalho, por exemplo, execute o código a seguir.

```
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
```

Se você tiver habilitado HTTPS com sucesso, o Livy enviará uma resposta indicando que o comando foi aceito e que o trabalho em lotes foi enviado.

```
{"id":1,"name":null,"owner":null,"proxyUser":null,"state":"starting","appId":null,"appInfo"
{"driverLogUrl":null,"sparkUiUrl":null},"log":["stdout: ", "\nstderr: ", "\nYARN
Diagnostics: "]}
```

Histórico de versões do Livy

A tabela a seguir lista a versão do Livy incluída em cada versão de lançamento da AmazonEMR, junto com os componentes instalados com o aplicativo. Para obter as versões dos componentes em cada versão, consulte a seção Versão de componentes da sua versão em [Versões de lançamento do Amazon EMR 7.x](#), [Versões de lançamento do Amazon EMR 6.x](#) ou [Versões de lançamento do Amazon EMR 5.x](#).

Informações das versões do Livy

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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-

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o Livy
emr-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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Livy	Componentes instalados com o 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

O Apache MXNet é uma biblioteca de aceleração projetada para criar redes neurais e outros aplicativos de aprendizado profundo. MXNet automatiza fluxos de trabalho comuns e otimiza os cálculos numéricos. MXNet ajuda você a projetar arquiteturas de redes neurais sem precisar se concentrar na implementação de cálculos de baixo nível, como operações de álgebra linear. MXNet está incluído na EMR versão 5.10.0 e posterior da Amazon.

Para obter mais informações, consulte o [MXNet site da Apache](#).

A tabela a seguir lista a versão MXNet incluída na versão mais recente da série Amazon EMR 7.x, junto com os componentes com os quais a Amazon é EMR instalada. MXNet

Para a versão dos componentes instalados MXNet nesta versão, consulte [Versões de componentes da versão 7.2.0](#).

MXNet informações sobre a versão do emr-7.2.0

Gravadora EMR de lançamento da Amazon	MXNet Versão	Componentes instalados com o 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

A tabela a seguir lista a versão MXNet incluída na versão mais recente da série EMR 6.x da Amazon, junto com os componentes com os quais a Amazon é EMR instalada. MXNet

Para a versão dos componentes instalados MXNet nesta versão, consulte Versões de [componentes da versão 6.15.0](#).

MXNet informações da versão do emr-6.15.0

Gravadora EMR de lançamento da Amazon	MXNet Versão	Componentes instalados com o 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

A tabela a seguir lista a versão MXNet incluída na versão mais recente da série Amazon EMR 5.x, junto com os componentes com os quais a Amazon é EMR instalada. MXNet

Para a versão dos componentes instalados MXNet nesta versão, consulte Versões de componentes da [versão 5.36.2](#).

MXNet informações da versão do emr-5.36.2

Gravadora EMR de lançamento da Amazon	MXNet Versão	Componentes instalados com o 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 histórico de lançamentos

A tabela a seguir lista a versão MXNet incluída em cada versão de lançamento da AmazonEMR, junto com os componentes instalados com o aplicativo. Para obter as versões dos componentes em cada versão, consulte a seção Versão de componentes da sua versão em [Versões de lançamento do Amazon EMR 7.x](#), [Versões de lançamento do Amazon EMR 6.x](#) ou [Versões de lançamento do Amazon EMR 5.x](#).

MXNet informações sobre a versão

Gravadora Amazon EMR Release	MXNet Versão	Componentes instalados com 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-

Gravadora Amazon EMR Release	MXNetVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	MXNetVersão	Componentes instalados com MXNet
emr-6.13.0	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-6.12.0	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-6.11.1	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-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Gravadora Amazon EMR Release	MXNetVersão	Componentes instalados com MXNet
emr-6.11.0	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-6.10.1	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-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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Gravadora Amazon EMR Release	MXNetVersão	Componentes instalados com 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Gravadora Amazon EMR Release	MXNetVersão	Componentes instalados com MXNet
emr-6.8.0	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-6.7.0	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-5.36.1	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

Gravadora Amazon EMR Release	MXNetVersão	Componentes instalados com MXNet
emr-5.36.0	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-resource-manager, 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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, 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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv

Gravadora Amazon EMR Release	MXNetVersão	Componentes instalados com MXNet
emr-6.5.0	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-6.4.0	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-6.3.1	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

Gravadora Amazon EMR Release	MXNetVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	MXNetVersão	Componentes instalados com MXNet
emr-6.1.1	1.6.0	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-6.1.0	1.6.0	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-6.0.1	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-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Gravadora Amazon EMR Release	MXNetVersão	Componentes instalados com 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-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, mxnet, opencv

Gravadora Amazon EMR Release	MXNetVersão	Componentes instalados com MXNet
emr-5.33.0	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-5.32.1	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-5.32.0	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

Gravadora Amazon EMR Release	MXNetVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	MXNetVersão	Componentes instalados com MXNet
emr-5.30.1	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-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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv
emr-5.29.0	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

Gravadora Amazon EMR Release	MXNetVersão	Componentes instalados com 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Gravadora Amazon EMR Release	MXNetVersão	Componentes instalados com MXNet
emr-5.27.0	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-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-https-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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Gravadora Amazon EMR Release	MXNetVersão	Componentes instalados com 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Gravadora Amazon EMR Release	MXNetVersão	Componentes instalados com MXNet
emr-5.23.0	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-5.22.0	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-5.21.2	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

Gravadora Amazon EMR Release	MXNetVersão	Componentes instalados com MXNet
emr-5.21.1	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-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-httpfs-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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet, opencv

Gravadora Amazon EMR Release	MXNetVersão	Componentes instalados com MXNet
emr-5.20.0	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-5.19.1	1.3.0	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-5.19.0	1.3.0	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

Gravadora Amazon EMR Release	MXNetVersão	Componentes instalados com MXNet
emr-5.18.1	1.2.0	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
emr-5.18.0	1.2.0	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
emr-5.17.2	1.2.0	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

Gravadora Amazon EMR Release	MXNetVersão	Componentes instalados com MXNet
emr-5.17.1	1.2.0	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-5.17.0	1.2.0	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-5.16.1	1.2.0	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

Gravadora Amazon EMR Release	MXNetVersão	Componentes instalados com MXNet
emr-5.16.0	1.2.0	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-5.15.1	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-5.15.0	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

Gravadora Amazon EMR Release	MXNetVersão	Componentes instalados com 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

Gravadora Amazon EMR Release	MXNetVersão	Componentes instalados com MXNet
emr-5.13.1	1.0.0	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-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-https-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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet

Gravadora Amazon EMR Release	MXNetVersão	Componentes instalados com MXNet
emr-5.12.2	1.0.0	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-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-https-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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet

Gravadora Amazon EMR Release	MXNetVersão	Componentes instalados com 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, mxnet

Gravadora Amazon EMR Release	MXNetVersão	Componentes instalados com MXNet
emr-5.11.1	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-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-https-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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mxnet

Gravadora Amazon EMR Release	MXNetVersão	Componentes instalados com 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

Use o programador de fluxo de trabalho do Apache Oozie para gerenciar e coordenar trabalhos do Hadoop. Para obter mais informações, consulte <http://oozie.apache.org/>.

A interface web nativa do Oozie não é compatível com a AmazonEMR. Para usar uma interface de front-end para o Oozie, tente o aplicativo Hue do Oozie. Para ter mais informações, consulte [Hue](#). O Oozie está incluído na EMR versão 5.0.0 e posterior da Amazon. Oozie é incluído como um aplicativo em sandbox em versões anteriores. Para ter mais informações, consulte [Versões de lançamento do Amazon EMR 4.x](#).

Se você usa um Amazon Linux personalizado AMI baseado em um Amazon Linux AMI com uma data de criação de 11/08/2018, o servidor Oozie falhará ao iniciar. Se você usa o Oozie, crie um personalizado AMI com base em um Amazon Linux AMI ID com uma data de criação diferente. Você pode usar o AWS CLI comando a seguir para retornar uma lista de imagens IDs para todo o HVM Amazon Linux AMIs com uma versão 2018.03, junto com a data de lançamento, para que você possa escolher um Amazon Linux apropriado AMI como sua base. MyRegion Substitua pelo seu identificador de região, como 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
```

A tabela a seguir lista a versão do Oozie incluída na versão mais recente da série Amazon EMR 7.x, junto com os componentes que a Amazon EMR instala com o Oozie.

Para a versão dos componentes instalados com o Oozie nesta versão, consulte [Versões de componentes da versão 7.2.0](#).

Informações sobre a versão do Oozie para o emr-7.2.0

Gravadora EMR de lançamento da Amazon	Versão do Oozie	Componentes instalados com o 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,

Gravadora EMR de lançamento da Amazon	Versão do Oozie	Componentes instalados com o Oozie
		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

A tabela a seguir lista a versão do Oozie incluída na versão mais recente da série EMR 6.x da Amazon, junto com os componentes que a Amazon EMR instala com o Oozie.

Para obter a versão dos componentes instalados com o Oozie nessa versão, consulte [Release 6.15.0 Component Versions](#).

Informações de versão do Oozie para o emr-6.15.0

Gravadora EMR de lançamento da Amazon	Versão do Oozie	Componentes instalados com o 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-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

A tabela a seguir lista a versão do Oozie incluída na versão mais recente da série Amazon EMR 5.x, junto com os componentes que a Amazon EMR instala com o Oozie.

Para a versão dos componentes instalados com o Oozie nesta versão, consulte Versões de componentes da [versão 5.36.2](#).

Informações sobre a versão do Oozie para o emr-5.36.2

Gravadora EMR de lançamento da Amazon	Versão do Oozie	Componentes instalados com o 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

Tópicos

- [Usando o Oozie com um banco de dados remoto na Amazon RDS](#)
- [Configurar a versão do Java para Oozie](#)
- [Histórico de versões do Oozie](#)

Usando o Oozie com um banco de dados remoto na Amazon RDS

Por padrão, as informações do usuário e os históricos de consultas do Oozie são armazenados em um SQL banco de dados My local no nó principal. Como alternativa, você pode criar um ou mais clusters habilitados para Oozie usando uma configuração armazenada no Amazon S3 e um banco de dados My no Amazon Relational Database Service (Amazon) RDS. Isso permite que você mantenha as informações do usuário e o histórico de consultas criado pela Oozie sem manter seu

EMR cluster da Amazon em execução. Recomendamos o uso da criptografia do lado do servidor do Amazon S3 para armazenar o arquivo de configuração.

Primeiro, crie o banco de dados remoto para o Oozie.

Para criar o Meu SQL banco de dados externo

1. Abra o RDS console da Amazon em <https://console.aws.amazon.com/rds/>.
2. Escolha Launch a DB Instance.
3. Escolha Meu SQL e, em seguida, selecione Selecionar.
4. Deixe a seleção padrão de Implantação Multi-AZ e IOPS Armazenamento Provisionado e escolha Avançar.
5. Deixe as especificações da instância em seus valores padrão, especifique configurações e selecione Next (Próximo).
6. Na página Configure Advanced Settings (Definir configurações avançadas), escolha um nome do grupo de segurança apropriado e o nome do banco de dados. O grupo de segurança que você usa deve pelo menos permitir TCP acesso de entrada para a porta 3306 a partir do nó principal do seu cluster. Se você não tiver criado o cluster neste ponto, poderá permitir que todos os hosts se conectem à porta 3306 e ajustar o security group depois de ter executado o cluster. Escolha Launch DB Instance (Iniciar instância de banco de dados).
7. No RDS Painel, selecione Instâncias e selecione a instância que você acabou de criar. Quando seu banco de dados estiver disponível, anote o nome do banco de dados, o nome de usuário, a senha e o nome do host da RDS instância. Você usará essas informações ao criar e configurar o cluster.

Para especificar um SQL banco de dados My externo para Oozie ao iniciar um cluster usando o AWS CLI

Para especificar um SQL banco de dados My externo para Oozie ao iniciar um cluster usando o AWS CLI, use as informações que você observou ao criar sua RDS instância para configuração `oozie-site` com um objeto de configuração.

Note

Você pode criar vários clusters que usam o mesmo banco de dados externo, mas cada cluster compartilhará informações de usuários e o histórico de usuários.

- Usando o AWS CLI, crie um cluster com o Oozie instalado, usando o banco de dados externo que você criou e referenciando um arquivo de configuração com uma classificação de configuração para o Oozie que especifica as propriedades do banco de dados. O exemplo a seguir cria um cluster com o Oozie instalado, fazendo referência a um arquivo de configuração no Amazon S3, `myConfig.json`, que especifica a configuração do banco de dados.

Note

Os caracteres de continuação de linha do Linux (`\`) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (`^`).

```
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
```

Os conteúdos de exemplo do arquivo `myConfig.json` são mostrados abaixo. Substituir *JDBC URL*, *username* e *password* com o JDBC URL nome de usuário e a senha da sua RDS instância.

Important

JDBCURLÉ necessário incluir o nome do banco de dados como sufixo. Por exemplo, `jdbc:mysql://.xxxxxxxxx.us-east-1.rds.amazonaws.com:3306/ dbname. oozie-external-db`

```
[{
  "Classification": "oozie-site",
  "Properties": {
    "oozie.service.JPIService.jdbc.driver": "org.mariadb.jdbc.Driver",
    "oozie.service.JPIService.jdbc.url": "JDBC URL",

    "oozie.service.JPIService.jdbc.username": "username",
    "oozie.service.JPIService.jdbc.password": "password"
  }
}]
```

```
    },  
    "Configurations": []  
  ]  
}
```

Configurar a versão do Java para Oozie

O Oozie executa vários processos Java Virtual Machine (JVM). Esta página explica como configurar a versão do Java para cada processo.

- **Servidor do Oozie:** defina `JAVA_HOME` na classificação `oozie-env` para atualizar a versão do Java para o `EmbeddedOozieServer`.
- **Oozie Launcher AM:** Oozie Launcher AM é um trabalho de MR (MapReduce) de mapeador único que invoca as bibliotecas de clientes de aplicações apropriadas, como Hadoop e Hive. A menos que seja configurado de outra forma, as versões de tempo de execução do Oozie Launcher AM são as mesmas dos tempos de execução Java do Hadoop no cluster. EMR Para configurar o runtime do Java para o Oozie Launcher AM, defina a seguinte propriedade no `workflow.xml` para o trabalho:

```
<property>  
  <name>mapred.child.env</name>  
  <value>JAVA_HOME=/path/to/JAVA_HOME</value>  
</property>
```

Essa propriedade garante que o Oozie Launcher AM para o trabalho do Oozie seja executado na versão do Java especificada por você, em vez de na versão do Java definida no Hadoop.

- **Executável do cliente da aplicação:** como o Oozie Launcher AM invoca o cliente da aplicação por padrão, o runtime do Java para o executável do cliente é o mesmo que o do Oozie Launcher AM.
- **Aplicativos iniciados por um trabalho do Oozie:** a menos que especificado de outra forma, as versões de tempo de execução do aplicativo real JVMs que são iniciadas por um trabalho do Oozie são as mesmas dos tempos de execução do Java para o Hadoop no cluster. EMR Dependendo do tipo de ação de fluxo de trabalho do Oozie usada para iniciar o aplicativo em uma tarefa do Oozie (ação do Spark ou do Hive), você pode atualizar o tempo de execução Java padrão para o aplicativo real JVMs na tarefa do `workflow.xml` Oozie.

Histórico de versões do Oozie

A tabela a seguir lista a versão do Oozie incluída em cada versão de lançamento da AmazonEMR, junto com os componentes instalados com o aplicativo. Para obter as versões dos componentes em cada versão, consulte a seção Versão de componentes da sua versão em [Versões de lançamento do Amazon EMR 7.x](#), [Versões de lançamento do Amazon EMR 6.x](#) ou [Versões de lançamento do Amazon EMR 5.x](#).

Informações de versão do Oozie

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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-

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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-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-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-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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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-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-6.11.0	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-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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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-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-6.10.0	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-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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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
emr-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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, oozie-client, oozie-server, tez-on-yarn

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Oozie	Componentes instalados com o 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

Notas de lançamento do Oozie por versão

Tópicos

- [Amazon EMR 7.1.0 - Mudanças no Oozie](#)

Amazon EMR 7.1.0 - Mudanças no Oozie

Tipo	Descrição
Upgrade	Atualiza o Netty para 4.1.100.Final.
Upgrade	Atualiza o Jetty para 9.4.53.v20231009.

Apache Phoenix

O Apache Phoenix é usado para OLTP análises operacionais, permitindo que você use SQL consultas padrão e trabalhe com uma JDBC APIs loja de apoio do ApacheHBase. Para obter mais informações, consulte [Phoenix em 15 minutos ou menos](#). O Phoenix está incluído na EMR versão 4.7.0 e posterior da Amazon.

Se você atualizar de uma versão anterior da Amazon EMR para a EMR versão 5.4.0 ou posterior da Amazon e usar indexação secundária, atualize os índices locais conforme descrito na documentação do [Apache](#) Phoenix. A Amazon EMR remove as configurações necessárias da `hbase-site` classificação, mas os índices precisam ser preenchidos novamente. O sistema oferece suporte a atualizações de índices online e offline. As atualizações online são o padrão, o que significa que os índices são preenchidos novamente durante a inicialização de clientes do Phoenix versão 4.8.0 ou posterior. Para especificar atualizações off-line, defina a `phoenix.client.localIndexUpgrade` configuração como falsa na `phoenix-site` classificação e, em seguida, no nó principal SSH a ser executada `psql [zookeeper] -1`.

A tabela a seguir lista a versão do Phoenix incluída na versão mais recente da série Amazon EMR 7.x, junto com os componentes que a Amazon EMR instala com o Phoenix.

Para a versão dos componentes instalados com o Phoenix nesta versão, consulte [Versões de componentes da versão 7.2.0](#).

Informações sobre a versão do Phoenix para o emr-7.2.0

Gravadora EMR de lançamento da Amazon	Versão do Phoenix	Componentes instalados com o 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-

Gravadora EMR de lançamento da Amazon	Versão do Phoenix	Componentes instalados com o Phoenix
		timeline-server, hbase-master, hbase-client, hbase-region-server, hbase-operator-tools, phoenix-library, phoenix-connectors, phoenix-query-server, zookeeper-client, zookeeper-server

A tabela a seguir lista a versão do Phoenix incluída na versão mais recente da série EMR 6.x da Amazon, junto com os componentes que a Amazon EMR instala com o Phoenix.

Para obter a versão dos componentes instalados com o Phoenix nessa versão, consulte [Release 6.15.0 Component Versions](#).

Informações de versão do Phoenix para o emr-6.15.0

Gravadora EMR de lançamento da Amazon	Versão do Phoenix	Componentes instalados com o 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-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-

Gravadora EMR de lançamento da Amazon	Versão do Phoenix	Componentes instalados com o Phoenix
		server, zookeeper-client, zookeeper-server

A tabela a seguir lista a versão do Phoenix incluída na versão mais recente da série Amazon EMR 5.x, junto com os componentes que a Amazon EMR instala com o Phoenix.

Para a versão dos componentes instalados com o Phoenix nesta versão, consulte Versões de componentes da [versão 5.36.2](#).

Informações sobre a versão do Phoenix para o emr-5.36.2

Gravadora EMR de lançamento da Amazon	Versão do Phoenix	Componentes instalados com o 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

Tópicos

- [Criar um cluster com o Phoenix](#)
- [Clientes do Phoenix](#)

- [Histórico de versões do Phoenix](#)

Criar um cluster com o Phoenix

Para instalar o Phoenix, escolha o aplicativo ao criar um cluster usando o console ou a AWS CLI. Os procedimentos e exemplos a seguir mostram como criar um cluster com Phoenix HBase e. Para obter mais informações sobre a criação de clusters usando o console, incluindo opções avançadas, consulte [Planejar e configurar clusters](#) no Amazon EMR Management Guide.

Para iniciar um cluster com o Phoenix instalado usando as Quick Options (Opções rápidas) para criar um cluster no console

1. Abra o EMR console da Amazon em <https://console.aws.amazon.com/emr>.
2. Escolha Create cluster (Criar cluster) para usar Quick Create.
3. Em Software Configuration (Configuração do software), escolha a versão mais recente apropriada para o seu aplicativo. O Phoenix aparece como uma opção somente quando a versão Amazon Release Version emr-4.7.0 ou posterior é selecionada.
4. Para Aplicativos, escolha a segunda opção HBase: HBase **ver** com Ganglia **ver**, Hadoop **ver**, Colmeia **ver**, Matiz **ver**, Fênix **ver**, e ZooKeeper **ver**.
5. Selecione outras opções conforme necessário e escolha Create cluster (Criar cluster).

Note

Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

O exemplo a seguir inicia um cluster com o Phoenix instalado usando as definições de configuração padrão.

Para iniciar um cluster com o Phoenix e HBase usar o AWS CLI

- Crie o cluster usando o seguinte comando:

```
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
```

Personalizar as configurações do Phoenix

Ao criar um cluster, você configurar o Phoenix definindo os valores em `hbase-site.xml` usando a classificação da configuração `hbase-site`.

Para obter mais informações, consulte [Configuração e ajustes](#) na documentação do Phoenix.

O exemplo a seguir demonstra o uso de um JSON arquivo armazenado no Amazon S3 para especificar o valor `false` da propriedade `phoenix.schema.dropMetaData`. Várias propriedades podem ser especificadas para uma única classificação. Para ter mais informações, consulte [Configurar aplicações](#). Em seguida, o `create-cluster` comando faz referência ao JSON arquivo como `--configurations` parâmetro.

O conteúdo do JSON arquivo salvo em `/mybucket/myfolder/myconfig.json` é o seguinte.

```
[  
  {  
    "Classification": "hbase-site",  
    "Properties": {  
      "phoenix.schema.dropMetaData": "false"  
    }  
  }  
]
```

O `create cluster` comando que faz referência ao JSON arquivo é mostrado no exemplo a seguir.

```
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

A solicitação de reconfiguração para qualquer classificação de configuração do Phoenix só é suportada na Amazon EMR versão 5.23.0 e posterior, e não é suportada na Amazon EMR versão 5.21.0 ou 5.22.0. Para obter mais informações, consulte [Fornecer uma configuração para um grupo de instâncias em um cluster em execução](#)

Clientes do Phoenix

Você se conecta ao Phoenix usando um JDBC cliente criado com dependências completas ou usando o “thin client” que usa o Phoenix Query Server e só pode ser executado em um nó principal de um cluster (por exemplo, usando um SQL cliente, uma etapa, linha de comando, encaminhamento de SSH porta etc.). Ao usar o JDBC cliente “fat”, ele ainda precisa ter acesso a todos os nós do cluster porque se conecta diretamente aos HBase serviços. O cliente “magro” do Phoenix precisa apenas do acesso ao Phoenix Query Server em uma porta padrão 8765. Há vários [scripts](#) no Phoenix que usam esses clientes.

Use uma EMR etapa da Amazon para consultar usando o Phoenix

O procedimento a seguir restaura um instantâneo HBase e usa esses dados para executar uma consulta Phoenix. Você pode estender este exemplo ou criar um novo script que utiliza os clientes do Phoenix para atender às suas necessidades.

1. Crie um cluster com o Phoenix instalado usando o seguinte comando:

```
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. Crie e, em seguida, carregue os seguintes arquivos para o 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/phoenix/bin/sqlline-thin.py http://localhost:8765 /home/hadoop/
phoenixQuery.sql
```

phoenixQuery.sql

 Note

Você só precisa incluir COLUMN_ENCODED_BYTES=0 no exemplo a seguir ao usar as EMR versões 5.26.0 e superiores da 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";
```

Use o AWS CLI para enviar os arquivos para o bucket do S3:

```
aws s3 cp copySnapshot.sh s3://myBucket/  
aws s3 cp runQuery.sh s3://myBucket/  
aws s3 cp phoenixQuery.sql s3://myBucket/
```

3. Crie uma tabela usando as seguintes etapas enviadas para o cluster que você criou na Etapa 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

```

- Use o `script-runner.jar` para executar o script `copySnapshot.sh` que foi anteriormente carregado para o bucket do 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"]

```

Isso executa um MapReduce trabalho para copiar seus dados de snapshot para o clusterHDFS.

- Restaure o snapshot que você copiou para o cluster usando a seguinte etapa:

`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

```

- Use o `script-runner.jar` para executar o script `runQuery.sh` que foi anteriormente carregado para o bucket do 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"]
```

A consulta é executada e obtém os resultados do stdout da etapa. Pode levar alguns minutos para a etapa ser concluída.

7. Inspecione os resultados das etapas stdout no registro URI que você usou ao criar o cluster na Etapa 1. Os resultados devem ser algo semelhante a:

```
+-----+-----+
|          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                            |
+-----+-----+
```

Histórico de versões do Phoenix

A tabela a seguir lista a versão do Phoenix incluída em cada versão de lançamento da AmazonEMR, junto com os componentes instalados com o aplicativo. Para obter as versões dos componentes em cada versão, consulte a seção Versão de componentes da sua versão em [Versões de lançamento do Amazon EMR 7.x](#), [Versões de lançamento do Amazon EMR 6.x](#) ou [Versões de lançamento do Amazon EMR 5.x](#).

Informações sobre a versão do Phoenix

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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-region-server, phoenix-library, phoenix-query-server,

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o Phoenix
		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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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-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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server
emr-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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, hbase-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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-master, 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-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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-hmaster, 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-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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-hmaster, 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-hmaster, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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-master, 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-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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-master, 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-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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-master, 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-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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-master, 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-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Gravadora Amazon EMR Release	Versão do Phoenix	Componentes instalados com o 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-master, 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-master, hbase-client, hbase-region-server, phoenix-library, phoenix-query-server, zookeeper-client, zookeeper-server

Apache Pig

O Apache Pig é uma biblioteca Apache de código aberto que é executada sobre o Hadoop, fornecendo uma linguagem de scripts que você pode usar para transformar grandes conjuntos de dados sem precisar gravar códigos complexos em uma linguagem de computação de nível inferior, como Java. A biblioteca usa comandos SQL semelhantes a comandos escritos em uma linguagem chamada Pig Latin e converte esses comandos em tarefas Tez com base em gráficos acíclicos direcionados () ou programas. DAGs MapReduce O Pig funciona com dados estruturados e não estruturados em vários formatos. Para obter mais informações sobre o Pig, consulte <http://pig.apache.org/>.

Você pode executar comandos do Pig interativamente ou no modo em lote. Para usar o Pig interativamente, crie uma SSH conexão com o nó principal e envie comandos usando o shell Grunt. Para usar o Pig no modo em lote, escreva os scripts do Pig, carregue-os no Amazon S3 e envie-os como etapas do cluster. Para obter mais informações sobre como enviar trabalhos para um cluster, consulte [Enviar trabalho para um cluster](#) no Amazon EMR Management Guide.

Ao usar o Pig para gravar a saída em uma HCatalog tabela no Amazon S3, desative a gravação direta EMR da Amazon definindo `mapred.output.direct.NativeS3FileSystem` as propriedades `mapred.output.direct.EmrFileSystem` e como `false` Para ter mais informações, consulte [Usando HCatalog](#). Dentro de um script do Pig, você pode usar os comandos SET `mapred.output.direct.NativeS3FileSystem false` e SET `mapred.output.direct.EmrFileSystem false`.

A tabela a seguir lista a versão do Pig incluída na versão mais recente da série Amazon EMR 7.x, junto com os componentes que a Amazon EMR instala com o Pig.

Para a versão dos componentes instalados com o Pig nesta versão, consulte [Versões de componentes da versão 7.2.0](#).

Informações sobre a versão do Pig para o emr-7.2.0

Gravadora EMR de lançamento da Amazon	Versão do Pig	Componentes instalados com o Pig
emr-7.2.0	Pig 0.17.0	emrfs, emr-ddb, emr-goodies, emr-kinesis, emr-s3-distcp, hadoop-client, hadoop-ma

Gravadora EMR de lançamento da Amazon	Versão do Pig	Componentes instalados com o Pig
		pred, hadoop-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

A tabela a seguir lista a versão do Pig incluída na versão mais recente da série Amazon EMR 6.x, junto com os componentes que a Amazon EMR instala com o Pig.

Para obter a versão dos componentes instalados com o Pig nessa versão, consulte [Release 6.15.0 Component Versions](#).

Informações de versão do Pig para o emr-6.15.0

Gravadora EMR de lançamento da Amazon	Versão do Pig	Componentes instalados com o 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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker

A tabela a seguir lista a versão do Pig incluída na versão mais recente da série Amazon EMR 5.x, junto com os componentes que a Amazon EMR instala com o Pig.

Para a versão dos componentes instalados com o Pig nesta versão, consulte Versões de componentes da [versão 5.36.2](#).

Informações sobre a versão do Pig para o emr-5.36.2

Gravadora EMR de lançamento da Amazon	Versão do Pig	Componentes instalados com o 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

Tópicos

- [Enviar um trabalho do Pig](#)
- [Chamar funções definidas pelo usuário no Pig](#)
- [Histórico de versões do Pig](#)

Enviar um trabalho do Pig

Esta seção demonstra o envio do trabalho do Pig para um cluster da Amazon EMR. Os exemplos a seguir geram um relatório contendo o total de bytes transferidos, uma lista dos 50 principais endereços IP, uma lista dos 50 principais indicadores externos e os 50 principais termos de pesquisa usando o Bing e o Google. O script do Pig está localizado no bucket `s3://elasticmapreduce/samples/pig-apache/do-reports2.pig` do Amazon S3. Os dados de entrada estão

localizados no bucket `s3://elasticmapreduce/samples/pig-apache/input` do Amazon S3. Os dados de saída são salvos em um bucket do Amazon S3.

Envie trabalhos do Pig usando o console da Amazon EMR

Este exemplo descreve como usar o EMR console da Amazon para adicionar uma etapa do Pig a um cluster.

Para enviar uma etapa do Pig

1. Abra o EMR console da Amazon em <https://console.aws.amazon.com/emr>.
2. Escolha Criar cluster para criar um cluster com o Pig instalado. Para ver as etapas sobre como criar um cluster, consulte [Planejar e configurar um EMR cluster da Amazon](#).
3. Abra um terminal e SSH entre no nó principal do seu cluster seguindo as etapas descritas em [Conecte-se ao nó principal usando SSH](#). Depois de fazer isso, execute as etapas a seguir.

```
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. No console, clique em Lista de clusters e selecione o nome do cluster que você criou.
5. Role até a seção Steps (Etapas) e expanda-a. Em seguida, escolha Add step (Adicionar etapa).
6. Na caixa de diálogo Add Step (Adicionar etapa):
 - Para Step type (Tipo de etapa), escolha Pig program (Programa Pig).
 - Para Name (Nome), aceite o nome padrão (programa Pig) ou digite um novo nome.
 - Para Script S3 location (Local do script do S3), digite o local do script do Pig. Por exemplo: **s3://elasticmapreduce/samples/pig-apache/do-reports2.pig**.
 - Para Input S3 location (Local de entrada do S3), digite o local dos dados de entrada. Por exemplo: **s3://elasticmapreduce/samples/pig-apache/input**.
 - Em Local de saída do S3, digite ou navegue até o nome do bucket de saída do Amazon S3.
 - Para Arguments (Argumentos), deixe o campo em branco.
 - Para Action on failure (Ação na falha), aceite a opção padrão Continue (Continuar).
7. Escolha Adicionar. A etapa é exibida no console com o status Pendente.
8. O status da etapa muda de Pending (Pendente) para Running (Em execução) e depois para Completed (Concluído) conforme ela é executada. Para atualizar o status, escolha o ícone

Refresh (Atualizar) acima da coluna Actions (Ações). Quando sua etapa estiver concluída, verifique o bucket do Amazon S3 para confirmar se os arquivos de saída da etapa do Pig estão lá.

Envie o trabalho do Pig usando o AWS CLI

Para enviar uma etapa do Pig usando o AWS CLI

Ao iniciar um cluster usando o AWS CLI, use o `--applications` parâmetro para instalar o Pig. Para enviar uma etapa do Pig, use o parâmetro `--steps`.

1. Para iniciar um cluster com o Pig instalado, digite o seguinte comando, substituindo *myKey* e *DOC-EXAMPLE-BUCKET/* com o nome do seu EC2 key pair e do bucket 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

Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

Quando você especifica a contagem de instâncias sem usar o parâmetro `--instance-groups`, um único nó principal é executado, e as instâncias restantes são executadas como nós core. Todos os nós usam o tipo de instância especificado no comando.

Note

Se você ainda não criou a função de EMR serviço e o perfil de EC2 instância padrão, digite `aws emr create-default-roles` para criá-los antes de digitar o `create-cluster` subcomando.

- Para enviar uma etapa do Pig, digite o seguinte comando, substituindo *myClusterId* e *DOC-EXAMPLE-BUCKET* com o ID do seu cluster e o nome do seu bucket do 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]
```

Esse comando retornará um ID de etapa, que você pode usar para verificar o State da sua etapa.

- Consulte o status da sua etapa com o comando `describe-step`.

```
aws emr describe-step --cluster-id myClusterId --step-id s-1XXXXXXXXXXA
```

O State da etapa é alterado de PENDING para RUNNING e para COMPLETED, conforme a etapa é executada. Quando sua etapa estiver concluída, verifique o bucket do Amazon S3 para confirmar se os arquivos de saída da etapa do Pig estão lá.

Para obter mais informações sobre o uso dos EMR comandos da Amazon no AWS CLI, consulte a [Referência de AWS CLI comandos](#).

Chamar funções definidas pelo usuário no Pig

O Pig fornece a capacidade de chamar funções definidas pelo usuário (UDFs) de dentro dos scripts do Pig. Você pode fazer isso para implementar o processamento personalizado em seus scripts do Pig. As linguagens atualmente suportadas são Java, Python/Jython e JavaScript (embora JavaScript o suporte ainda seja experimental).

As seções a seguir descrevem como registrar suas funções no Pig para que você possa chamá-las a partir do shell ou de scripts do Pig. Para obter mais informações sobre como usar UDFs com o Pig, consulte a [documentação do Pig](#) para sua versão do Pig.

Chame JAR arquivos do Pig

Você pode usar JAR arquivos personalizados com o Pig usando o REGISTER comando em seu script do Pig. O JAR arquivo é local ou um sistema de arquivos remoto, como o Amazon S3. Quando o script Pig é executado, a Amazon EMR baixa o JAR arquivo automaticamente para o nó principal e, em seguida, carrega o JAR arquivo no cache distribuído do Hadoop. Dessa forma, o JAR arquivo é usado automaticamente, conforme necessário, por todas as instâncias no cluster.

Para usar JAR arquivos com o Pig

1. Faça o upload do seu JAR arquivo personalizado no Amazon S3.
2. Use o REGISTER comando em seu script Pig para especificar o bucket no Amazon S3 do JAR arquivo personalizado.

```
REGISTER s3://mybucket/path/mycustomjar.jar;
```

Chamar scripts em Python/Jython no Pig

Você pode registrar scripts em Python no Pig e, em seguida, chamar funções nesses scripts a partir do shell do Pig ou de um script do Pig. Para fazer isso, especifique o local do script com a palavra-chave `register`.

Como o Pig em escrito em Java, ele usa o mecanismo de script Jython para analisar scripts do Python. Para obter mais informações sobre Jython, acesse <http://www.jython.org/>.

Para chamar scripts em Python/Jython a partir do Pig

1. Escreva um script em Python e carregue-o para um local no Amazon S3. Esse deve ser um bucket de propriedade da mesma conta que estiver criando o cluster do Pig, ou um bucket que tenha permissões definidas de tal forma que a conta que criou o cluster pode acessá-lo. Neste exemplo, o upload do script é feito para `s3://mybucket/pig/python`.
2. Inicie um cluster Pig. Se você estiver acessando o Pig a partir do shell Grunt, execute um cluster interativo. Se você estiver executando comandos do Pig em um script, inicie um cluster Pig com

script. Este exemplo inicia um cluster interativo. Para obter mais informações sobre como criar um cluster Pig, consulte [Enviar um trabalho do Pig](#).

3. Para um cluster interativo, use SSH para se conectar ao nó principal e executar o shell Grunt. Para obter mais informações, consulte [SSH no nó principal](#).
4. Execute o shell Grunt do Pig, digitando `pig` na linha de comando:

```
pig
```

5. Registre a biblioteca Jython e o script do Python no Pig usando a palavra-chave `register` no prompt de comando do Grunt, como mostrado no comando a seguir, no qual você deve especificar o local do seu script no Amazon S3:

```
grunt> register 'lib/jython.jar';  
grunt> register 's3://mybucket/pig/python/myscript.py' using jython as myfunctions;
```

6. Carregue os dados de entrada. O exemplo a seguir carrega a entrada de um local do Amazon S3:

```
grunt> input = load 's3://mybucket/input/data.txt' using TextLoader as  
(line:chararray);
```

7. Agora, você pode chamar funções no seu script dentro Pig, fazendo referência a eles usando `myfunctions`:

```
grunt> output=foreach input generate myfunctions.myfunction($1);
```

Histórico de versões do Pig

A tabela a seguir lista a versão do Pig incluída em cada versão de lançamento da AmazonEMR, junto com os componentes instalados com o aplicativo. Para obter as versões dos componentes em cada versão, consulte a seção Versão de componentes da sua versão em [Versões de lançamento do Amazon EMR 7.x](#), [Versões de lançamento do Amazon EMR 6.x](#) ou [Versões de lançamento do Amazon EMR 5.x](#).

Informações de versão do Pig

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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-

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn, tez-on-worker

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o Pig
emr-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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, pig-client, tez-on-yarn

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, pig-client

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, pig-client

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, pig-client

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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-httplibs-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-httplibs-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-httplibs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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-httplibs-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-httplibs-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-httplibs-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client

Gravadora Amazon EMR Release	Versão do Pig	Componentes instalados com o 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-https-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, pig-client

Presto e Trino

Note

Presto SQL foi renomeado para Trino em dezembro de 2020. EMRAs versões 6.4.0 e posteriores da Amazon usam o nome Trino, enquanto as versões anteriores usam o nome Presto. SQL

O [Presto](#) é um mecanismo de SQL consulta rápido projetado para consultas analíticas interativas em grandes conjuntos de dados de várias fontes. Para obter mais informações, consulte o [Site do Presto](#). O Presto está incluído nas EMR versões 5.0.0 e posteriores da Amazon. As versões anteriores incluem o Presto como um aplicação em sandbox. Para obter mais informações, consulte [Versões de lançamento do Amazon EMR 4.x](#). As EMR versões 6.1.0 e posteriores da Amazon oferecem suporte ao [Trino](#) (PrestoSQL), além do Presto. Para obter mais informações, consulte [Instalação do PrestoDB e do Trino](#).

A tabela a seguir lista a versão do Presto incluída na versão mais recente da série Amazon EMR 7.x, junto com os componentes que a Amazon EMR instala com o Presto.

Para a versão dos componentes instalados com o Presto nesta versão, consulte [Versões de componentes da versão 7.2.0](#).

Informações sobre a versão do Presto para o emr-7.2.0

Gravadora EMR de lançamento da Amazon	Versão do Presto	Componentes instalados com o 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,

Gravadora EMR de lançamento da Amazon	Versão do Presto	Componentes instalados com o Presto
		hcatalog-server, mariadb-server, presto-coordinator, presto-worker

A tabela a seguir lista a versão do Presto incluída na versão mais recente da série EMR 6.x da Amazon, junto com os componentes que a Amazon EMR instala com o Presto.

Para obter a versão dos componentes instalados com o Presto nessa versão, consulte [Release 6.15.0 Component Versions](#).

Informações de versão do Presto para o emr-6.15.0

Gravadora EMR de lançamento da Amazon	Versão do Presto	Componentes instalados com o 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

A tabela a seguir lista a versão do Presto incluída na versão mais recente da série Amazon EMR 5.x, junto com os componentes que a Amazon EMR instala com o Presto.

Para a versão dos componentes instalados com o Presto nesta versão, consulte [Versões de componentes da versão 5.36.2](#).

Informações sobre a versão do Presto para o emr-5.36.2

Gravadora EMR de lançamento da Amazon	Versão do Presto	Componentes instalados com o 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

A tabela a seguir lista a versão do Trino (PrestoSQL) incluída na versão mais recente da série EMR 6.x da Amazon, junto com os componentes que a Amazon EMR instala com o Trino (Presto). SQL

Para a versão dos componentes instalados com o Trino (PrestoSQL) nesta versão, consulte Versões de componentes da [versão 6.15.0](#).

Informações sobre a versão do Trino (PrestoSQL) para o emr-6.15.0

Gravadora EMR de lançamento da Amazon	Versão Trino (PrestoSQL)	Componentes instalados com Trino (SQLPresto)
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-

Gravadora EMR de lançamento da Amazon	Versão Trino (PrestoSQL)	Componentes instalados com Trino (SQLPresto)
		server, mariadb-server, trino-coordinator, trino-worker

Tópicos

- [Usando o Presto com o AWS Glue Data Catalog](#)
- [Usar S3 Select Pushdown com o Presto para melhorar a performance](#)
- [Adicionar conectores de banco de dados](#)
- [UsandoSSL/TLS configurando LDAPS com o Presto na Amazon EMR](#)
- [Ativar o modo estrito do Presto](#)
- [Lidar com a perda de instâncias spot no Presto](#)
- [Execução tolerante a falhas no Trino](#)
- [Usar a escalabilidade automática do Presto com desativação tranquila](#)
- [Considerações sobre o Presto na Amazon EMR](#)
- [Histórico de versões do Presto](#)

Usando o Presto com o AWS Glue Data Catalog

Usando a EMR versão 5.10.0 da Amazon e posterior, você pode especificar o AWS Glue Data Catalog como a metastore padrão do Hive para o Presto. Recomendamos essa configuração quando você precisa de um metastore persistente ou de um metastore compartilhado por diferentes clusters, serviços, aplicações ou Contas da AWS.

AWS O Glue é um serviço totalmente gerenciado de extração, transformação e carregamento (ETL) que torna simples e econômico categorizar seus dados, limpá-los, enriquecê-los e movê-los de forma confiável entre vários armazenamentos de dados. O AWS Glue Data Catalog fornece um repositório unificado de metadados em uma variedade de fontes e formatos de dados, com integração com a Amazon, Amazon EMR Redshift, RDS Redshift Spectrum, Athena e qualquer aplicativo compatível com o metastore Apache Hive. AWS Os rastreadores do Glue podem inferir automaticamente o esquema dos dados de origem no Amazon S3 e armazenar os metadados associados no catálogo de dados. Para obter mais informações sobre o Catálogo de Dados, consulte [Preenchendo o Catálogo de Dados do AWS Glue](#) no Guia do Desenvolvedor do AWS Glue.

Taxas separadas se aplicam ao AWS Glue. Há uma taxa mensal para armazenar e acessar os metadados no Catálogo de Dados, uma taxa horária cobrada por minuto pelas ETL tarefas do AWS Glue e pelo tempo de execução do rastreador, e uma taxa horária cobrada por minuto para cada endpoint de desenvolvimento provisionado. O Catálogo de Dados permite o armazenamento de até um milhão de objetos gratuitamente. Se você armazenar mais de um milhão de objetos, será cobrado USD \$1 por cada 100.000 objetos acima de um milhão. Um objeto no Catálogo de Dados é uma tabela, uma partição ou um banco de dados. Para obter mais informações, consulte [Definição de preço do Glue](#).

Important

Se você criou tabelas usando o Amazon Athena ou o Amazon Redshift Spectrum antes de 14 de agosto de 2017, bancos de dados e tabelas são armazenados em um catálogo gerenciado pelo Athena, que é separado do Glue Data Catalog. Para integrar a Amazon EMR a essas tabelas, você deve fazer o upgrade para o AWS Glue Data Catalog. Para obter mais informações, consulte [Atualização para o catálogo de dados AWS Glue no Guia](#) do usuário do Amazon Athena.

Especificando o AWS Glue Data Catalog como metastore

Você pode especificar o AWS Glue Data Catalog como metastore usando o AWS Management Console, AWS CLI, ou Amazon EMR API. Ao usar o CLI ou API, você usa a classificação de configuração do Presto para especificar o Catálogo de Dados. Além disso, com o Amazon EMR 5.16.0 e versões posteriores, você pode usar a classificação de configuração para especificar um catálogo de dados em outro. No console da AWS, você pode especificar o Catálogo de Dados usando Opções avançadas ou Opções rápidas.

Console

Para especificar o AWS Glue Data Catalog como o metastore do Hive com o novo console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em <https://console.aws.amazon.com/emr>.
2. No painel de navegação esquerdo, escolha Clusters e, em seguida, escolha Criar cluster.
3. Em Pacote de aplicações, escolha Presto.

4. Em Configurações do Catálogo de Dados do AWS Glue, marque a caixa de seleção Usar para metadados de tabelas do Presto.
5. Escolha qualquer outra opção que se aplique ao cluster.
6. Para iniciar o cluster, escolha Criar cluster.

CLI

Para especificar o AWS Glue Data Catalog como o metastore padrão do Hive usando o AWS CLI

Para obter exemplos de como especificar as classificações de configuração a seguir ao criar um cluster, consulte [Configurar aplicações](#).

Amazon EMR 5.16.0 e versões posteriores

- Defina a `hive.metastore` propriedade `glue` como mostrado no JSON exemplo a seguir.

```
[
  {
    "Classification": "presto-connector-hive",
    "Properties": {
      "hive.metastore": "glue"
    }
  }
]
```

Para especificar um catálogo de dados em outro Conta da AWS, adicione a `hive.metastore.glue.catalogid` propriedade conforme mostrado no JSON exemplo a seguir. Substitua *acct-id* pela conta da AWS do Catálogo de Dados. O uso de um catálogo de dados em outro não Conta da AWS está disponível usando a Amazon EMR versão 5.15.0 e anteriores.

```
[
  {
    "Classification": "presto-connector-hive",
    "Properties": {
      "hive.metastore": "glue",
      "hive.metastore.glue.catalogid": "acct-id"
    }
  }
]
```

Amazon EMR 5.10.0 a 5.15.0

Defina a `hive.metastore.glue.datacatalog.enabled` propriedade como `true`, conforme mostrado no JSON exemplo a seguir:

```
[
  {
    "Classification": "presto-connector-hive",
    "Properties": {
      "hive.metastore.glue.datacatalog.enabled": "true"
    }
  }
]
```

Amazon EMR 6.1.0 e versões posteriores usando Presto SQL (Trino)

A partir da EMR versão 6.1.0, o Presto SQL também oferece suporte ao Glue como metastore padrão do Hive. Use a classificação `prestoql-connector-hive` de configuração e defina a `hive.metastore` propriedade como `glue`, conforme mostrado no JSON exemplo a seguir.

EMRAs versões 6.4.0 e posteriores da Amazon usam o novo nome Trino em vez de Presto. SQL Se você usa o Trino, substitua *prestoql-connector-hive* na seguinte classificação de configuração por `trino-connector-hive`.

```
[
  {
    "Classification": "prestoql-connector-hive",
    "Properties": {
      "hive.metastore": "glue"
    }
  }
]
```

Para alternar entre metastores em um cluster de execução prolongada, você pode definir esses valores manualmente, conforme apropriado para a sua versão, conectando-se ao nó principal, editando os valores de propriedade diretamente no arquivo `/etc/presto/conf/catalog/hive.properties` e reiniciando o servidor Presto (`sudo restart presto-server`).

Se você usar esse método com o Amazon EMR 5.15.0 e versões anteriores, verifique se `hive.table-statistics-enabled` está definido como `false`. Essa configuração não é necessária ao usar versões 5.16.0 e posteriores. No entanto, as estatísticas de tabela e de partição não são compatíveis.

IAMpermissões

O perfil da EC2 instância de um cluster deve ter IAM permissões para as ações do AWS Glue. Além disso, se você habilitar a criptografia para objetos do AWS Glue Data Catalog, a função também deverá ter permissão para criptografar, descriptografar e gerar o AWS KMS key usado para criptografia.

Permissões para ações do AWS Glue

Se você usa o perfil de EC2 instância padrão da AmazonEMR, nenhuma ação é necessária. A política `AmazonElasticMapReduceforEC2Role` gerenciada anexada ao `EMR_EC2_DefaultRole` permite todas as ações necessárias do AWS Glue. No entanto, se você especificar um perfil de EC2 instância e permissões personalizados, deverá configurar as ações apropriadas do AWS Glue. Use a política gerenciada `AmazonElasticMapReduceforEC2Role` como ponto de partida. Para obter mais informações, consulte [Função de serviço para EC2 instâncias de cluster \(perfil de EC2 instância\)](#) no Amazon EMR Management Guide.

Permissões para criptografar e descriptografar o Glue Data Catalog AWS

O perfil de instância precisa de permissão para criptografar e descriptografar dados usando a chave. Você não precisa configurar essas permissões se as duas instruções a seguir se aplicarem:

- Você ativa a criptografia para objetos do AWS Glue Data Catalog usando chaves gerenciadas para o AWS Glue.
- Você usa um cluster que está no mesmo catálogo Conta da AWS de dados do AWS Glue.

Caso contrário, você deverá adicionar a seguinte declaração à política de permissões anexada ao seu perfil de EC2 instância.

```
[
  {
    "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"
    }
  ]
}
]

```

Para obter mais informações sobre a criptografia do AWS Glue Data Catalog, consulte [Criptografando seu catálogo de dados](#) no AWS Glue Developer Guide.

Permissões baseadas em recursos

Se você usa o AWS Glue em conjunto com o Hive, o Spark ou o Presto na AmazonEMR, o AWS Glue oferece suporte a políticas baseadas em recursos para controlar o acesso aos recursos do Catálogo de Dados. Esses recursos incluem bancos de dados, tabelas, conexões e funções definidas pelo usuário. Para obter mais informações, consulte [Políticas baseadas em recursos no AWS Glue](#) no Guia do desenvolvedor do AWS Glue.

Ao usar políticas baseadas em recursos para limitar o acesso ao AWS Glue na AmazonEMR, o principal que você especifica na política de permissões deve ser a função ARN associada ao perfil de EC2 instância especificado quando um cluster é criado. Por exemplo, para uma política baseada em recursos anexada a um catálogo, você pode especificar a função da função de serviço padrão ARN para instâncias de clusterEC2, *EMR_EC2_DefaultRole* como oPrincipal, usando o formato mostrado no exemplo a seguir:

```
arn:aws:iam::acct-id:role/EMR_EC2_DefaultRole
```

A ferramenta *acct-id* pode ser diferente do ID da conta AWS Glue. Isso permite o acesso a partir de EMR clusters em contas diferentes. Você pode especificar várias entidades principais, cada uma de uma conta diferente.

Considerações ao usar o Catálogo de Dados do AWS Glue

Considere os seguintes itens ao usar o AWS Glue Data Catalog como metastore com o Presto:

- Não há suporte para renomear tabelas de dentro do AWS Glue.
- Quando você cria uma tabela do Hive sem especificar um LOCATION, os dados da tabela são armazenados no local especificado pela propriedade `hive.metastore.warehouse.dir`. Por padrão, esse é um local em HDFS. Se outro cluster precisar acessar a tabela, ele apresentará falha, a menos que tenha permissões adequadas para o cluster que criou a tabela. Além disso, como o HDFS armazenamento é transitório, se o cluster for encerrado, os dados da tabela serão perdidos e a tabela deverá ser recriada. Recomendamos que você especifique a LOCATION no Amazon S3 ao criar uma tabela do Hive usando o Glue. Como alternativa, você pode usar a classificação de configuração `hive-site` para especificar um local no Amazon S3 para `hive.metastore.warehouse.dir`, que se aplica a todas as tabelas do Hive. Se uma tabela for criada em um HDFS local e o cluster que a criou ainda estiver em execução, você poderá atualizar a localização da tabela para o Amazon S3 a partir do AWS Glue. Para obter mais informações, consulte Como [trabalhar com tabelas no console do AWS Glue](#) no Guia do desenvolvedor do AWS Glue.
- Valores de partição contendo aspas e apóstrofes não são compatíveis, por exemplo, `PARTITION (owner="Doe 's")`.
- [As estatísticas de coluna](#) não têm suporte para o emr-5.31.0 e versões posteriores.
- O uso da [autorização do Hive](#) não é compatível. Como alternativa, considere usar as [Políticas baseadas em recursos do AWS Glue](#). Para obter mais informações, consulte [Usar políticas baseadas em recursos para o catálogo de dados Amazon EMR Access to AWS Glue](#).

Usar S3 Select Pushdown com o Presto para melhorar a performance

Com a EMR versão 5.18.0 e posterior da Amazon, você pode usar o [S3 select](#) Pushdown com o Presto na Amazon. Esse atributo permite que o Presto “empurre” o trabalho de computação de operações de projeção (por exemplo, SELECT) e operações de predicado (por exemplo, WHERE) para o Amazon S3. Isso permite que as consultas recuperem somente os dados necessários do Amazon S3, o que pode melhorar o desempenho e reduzir a quantidade de dados transferidos entre a Amazon e o EMR Amazon S3 em alguns aplicativos.

O S3 Select Pushdown é adequado para minha aplicação?

Recomendamos que você avalie seus aplicativos com e sem o S3 Seleção Pushdown para ver se o uso pode ser adequado para o aplicativo.

Use as seguintes diretrizes para determinar se o seu aplicativo é adequado para o uso do S3 Select:

- Sua consulta filtra mais de metade do conjunto de dados original.
- Os predicados do filtro de consulta usam colunas que têm um tipo de dados compatível com o Presto e o S3 Select. Os tipos de dados real e duplo e o carimbo de data e hora não têm suporte no S3 Select Pushdown. É recomendável usar o tipo de dados decimal para dados numéricos. Para obter mais informações sobre tipos de dados compatíveis com o S3 Select, consulte [Tipos de dados](#) no Guia do usuário do Amazon Simple Storage Service.
- Sua conexão de rede entre o Amazon S3 e o EMR cluster da Amazon tem boa velocidade de transferência e largura de banda disponível. O Amazon S3 não compacta HTTP respostas, então é provável que o tamanho da resposta aumente para arquivos de entrada compactados.

Considerações e limitações

- Somente objetos armazenados em CSV formato são suportados. Os objetos podem ser descompactados ou compactados com gzip ou bzip2.
- A propriedade `AllowQuotedRecordDelimiters` não é compatível. Se essa propriedade for especificada, a consulta falhará.
- A criptografia do lado do servidor do Amazon S3 com chaves de criptografia fornecidas pelo cliente (SSE-C) e a criptografia do lado do cliente não são suportadas.
- O S3 Select Pushdown não substitui o uso de formatos de arquivo colunares ou compactados, como o Parquet. ORC

Habilitar o S3 Select Pushdown com o PrestoDB ou Trino

Para habilitar o S3 Select Pushdown para PrestoDB na AmazonEMR, use a classificação de `presto-connector-hive` configuração `hive.s3select-pushdown.enabled` para definir `true` como mostrado no exemplo abaixo. Para obter mais informações, consulte [Configurar aplicações](#). O valor `hive.s3select-pushdown.max-connections` também deve ser definido. Para a maioria dos aplicativos, a configuração padrão de `500` deve ser adequada. Para obter mais informações, consulte [Noções básicas e ajuste do `hive.s3select-pushdown.max-connections`](#) abaixo.

Para o Presto SQL nas EMR versões 6.1.0 a 6.3.0, substitua `presto-connector-hive` no exemplo abaixo por `presto-sql-connector-hive`

EMRAs versões 6.4.0 e posteriores da Amazon usam o novo nome Trino em vez de Presto. SQL Se você usa o Trino, substitua *presto-connector-hive* no exemplo abaixo por *trino-connector-hive*

```
[
  {
    "classification": "presto-connector-hive",
    "properties": {
      "hive.s3select-pushdown.enabled": "true",
      "hive.s3select-pushdown.max-connections": "500"
    }
  }
]
```

Noções básicas e ajuste do `hive.s3select-pushdown.max-connections`

Por padrão, o Presto usa EMRFS como seu sistema de arquivos. A configuração `fs.s3.maxConnections` na classificação de `emrfs-site` configuração específica o máximo permitido de conexões de clientes com o Amazon S3 EMRFS por meio do For Presto. Por padrão, é 500. O S3 Select Pushdown ignora EMRFS ao acessar o Amazon S3 para operações de predicados. Nesse caso, o valor de `hive.s3select-pushdown.max-connections` determina o número máximo de conexões de cliente permitidas para essas operações de nós do operador. No entanto, todas as solicitações para o Amazon S3 iniciadas pelo Presto que não sejam canceladas — por exemplo, GET operações — continuam sendo regidas pelo valor de `fs.s3.maxConnections`

Se o seu aplicativo tiver o erro "Tempo limite atingido em espera para conexão do grupo", aumente o valor de `hive.s3select-pushdown.max-connections` e `fs.s3.maxConnections`.

Adicionar conectores de banco de dados

Você pode usar classificações de configuração para configurar as propriedades JDBC do conector ao criar um cluster. As classificações de configuração começam com `presto-connector`, por exemplo, `presto-connector-postgresql`. As classificações de configuração disponíveis dependem da versão de EMR lançamento da Amazon. Para obter as classificações de configuração disponíveis com a versão de lançamento mais recente, consulte a Amazon [the section called “5.36.2 classificações de configuração”](#) EMR 5.36.2. Se você estiver usando uma versão diferente da AmazonEMR, consulte [Versões de lançamento do Amazon EMR 5.x](#) as classificações de configuração. Para obter mais informações sobre as propriedades que podem ser configuradas com cada conector, consulte <https://prestodb.io/docs/current/connector.html>.

Example —configurando um cluster com o conector Postgre SQL JDBC

Para iniciar um cluster com o SQL conector Postgre instalado e configurado, primeiro crie um JSON arquivo que especifique a classificação da configuração — por exemplo, `myConfig.json` — com o conteúdo a seguir e salve-o localmente.

Substitua as propriedades de conexão conforme apropriado para sua configuração e conforme mostrado no tópico do [SQLconector Postgre](#) na documentação do Presto.

```
[
  {
    "Classification": "presto-connector-postgresql",
    "Properties": {
      "connection-url": "jdbc:postgresql://example.net:5432/database",
      "connection-user": "MYUSER",
      "connection-password": "MYPASS"
    },
    "Configurations": []
  }
]
```

Ao criar o cluster, faça referência ao caminho para o JSON arquivo usando a `--configurations` opção, conforme mostrado no exemplo a seguir, onde `myConfig.json` está no mesmo diretório em que você executa o comando:

```
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
```

UsandoSSL/TLS e configurando LDAPS com o Presto na Amazon EMR

Com a EMR versão 5.6.0 e posterior da Amazon, você pode habilitarSSL/TLS para ajudar a [proteger a comunicação interna](#) entre os nós do Presto. Você pode fazer isso definindo uma configuração de segurança para criptografia em trânsito. Para obter mais informações, consulte [Opções de criptografia](#) e [Usar configurações de segurança para configurar a segurança do cluster](#) no Amazon EMR Management Guide.

Quando você usa uma configuração de segurança com criptografia em trânsito, a Amazon EMR faz o seguinte para o Presto:

- Distribui os artefatos ou certificados de criptografia que você especifica para a criptografia em trânsito em todo o cluster do Presto. Para obter mais informações, consulte [Fornecer certificados para criptografia de dados em trânsito](#).
- Define as seguintes propriedades usando a classificação de configuração `presto-config`, que corresponde ao arquivo `config.properties` para o Presto:
 - Define `http-server.http.enabled` como `false` em todos os nós, o que desativa HTTP em favor de HTTPS. Isso exige que você forneça certificados que funcionem para públicos e privados DNS ao definir a configuração de segurança para criptografia em trânsito. Uma maneira de fazer isso é usar certificados SAN (Nome alternativo do assunto) que suportam vários domínios.
 - Define os valores `http-server.https.*`. Para obter detalhes de configuração, consulte a [LDAPautenticação](#) na documentação do Presto.
- Para o Presto SQL (Trino) na EMR versão 6.1.0 e posterior, a Amazon configura EMR automaticamente uma chave secreta compartilhada para comunicação interna segura entre os nós do cluster. Você não precisa fazer qualquer configuração adicional para habilitar esse atributo de segurança e pode substituir a configuração com sua própria chave secreta. Para obter informações sobre a autenticação interna do Trino, consulte a [documentação do Trino 353: Proteger a comunicação interna](#).

Além disso, com a EMR versão 5.10.0 e posterior da Amazon, você pode configurar a [LDAPautenticação](#) para conexões de clientes com o coordenador do Presto usando HTTPS. Essa configuração usa secure LDAP (LDAPS). TLS deve estar habilitado em seu LDAP servidor, e o cluster Presto deve usar uma configuração de segurança com criptografia de dados em trânsito ativada. Configurações adicionais são necessárias. As opções de configuração são diferentes dependendo da versão de lançamento da Amazon EMR que você usa. Para obter mais informações, consulte [Usando a LDAP autenticação para o Presto na Amazon EMR](#).

O Presto na Amazon EMR usa a porta 8446 como interna HTTPS por padrão. A porta usada para comunicação interna deve ser a mesma usada para HTTPS acesso do cliente ao coordenador do Presto. A propriedade `http-server.https.port` na classificação de configuração `presto-config` especifica a porta.

Usando a LDAP autenticação para o Presto na Amazon EMR

Siga as etapas nesta seção para configurarLDAP. Consulte cada etapa para obter exemplos e links para mais informações.

Etapas para configurar a LDAP autenticação

- [Etapa 1: Reúna informações sobre seu LDAP servidor e copie o certificado do servidor para o Amazon S3](#)
- [Etapa 2: definir uma configuração de segurança](#)
- [Etapa 3: criar uma configuração JSON com as propriedades do Presto para LDAP](#)
- [Etapa 4: Crie o script para copiar o certificado LDAP do servidor e carregá-lo no Amazon S3](#)
- [Etapa 5: criar um cluster](#)

Etapa 1: Reúna informações sobre seu LDAP servidor e copie o certificado do servidor para o Amazon S3

Você precisará das informações e dos itens do seu LDAP servidor na seção a seguir para configurar a LDAP autenticação.

O endereço IP ou nome do host do LDAP servidor

O coordenador do Presto no nó EMR principal da Amazon deve ser capaz de acessar o LDAP servidor no endereço IP ou nome do host especificado. Por padrão, o Presto se comunica com o LDAP servidor usando a porta LDAPS 636. Se sua LDAP implementação exigir uma porta personalizada, você pode especificá-la usando a `ldap.url` propriedade com o Amazon EMR 5.16.0 ou posterior, ou usando `authentication.ldap.url` com versões anteriores. Substitua a porta personalizada em 636, conforme mostrado nos exemplos da classificação de configuração `presto-config` em [Etapa 3: criar uma configuração JSON com as propriedades do Presto para LDAP](#).

Verifique se todos os firewalls e grupos de segurança permitem o tráfego de entrada e saída na porta 636 (ou na porta personalizada) e também na porta 8446 (ou na porta personalizada), que é usada para comunicações do cluster interno.

O certificado LDAP do servidor

Você deve carregar o arquivo do certificado em um local seguro no Amazon S3. Para obter mais informações, consulte [Como carregar arquivos e pastas em um bucket do S3](#) no Guia do usuário do Amazon Simple Storage Service. Crie uma ação de bootstrap que copie esse certificado do

Amazon S3 em cada nó do cluster quando esse cluster é iniciado. Em [Etapa 4: Crie o script para copiar o certificado LDAP do servidor e carregá-lo no Amazon S3](#). O certificado de exemplo é `s3://MyBucket/ldap_server.crt`.

As configurações LDAP do servidor para vinculação anônima

Se a vinculação anônima estiver desativada no PrestoDB, você precisará do ID de usuário UID () e da senha de uma conta com permissões para se vincular ao servidor para que LDAP o servidor PrestoDB possa estabelecer uma conexão. Você especifica a senha UID e usando as `internal-communication.ldap.password` propriedades `internal-communication.ldap.user` e na classificação `presto-config` de configuração. O Amazon EMR 5.10.0 não oferece suporte a essas configurações, portanto, a vinculação anônima deve ser suportada no LDAP servidor quando você usa esta versão de lançamento.

Observe que o Trino não exige a configuração de vinculação anônima.

Para obter o status de vinculação anônima no LDAP servidor

- Use o comando [ldapwhoami](#) do cliente Linux, conforme mostrado no exemplo a seguir:

```
ldapwhoami -x -H ldaps://LDAPServerHostNameOrIPAddress
```

Se a associação anônima não for permitida, o comando retornará o seguinte:

```
ldap_bind: Inappropriate authentication (48)
additional info: anonymous bind disallowed
```

Para verificar se uma conta tem permissões para um LDAP servidor que usa autenticação simples

- Use o comando [ldapwhoami](#) do cliente Linux, conforme mostrado no exemplo a seguir. O exemplo usa um usuário fictício, `presto`, armazenado em um LDAP servidor Open executado em uma EC2 instância com o nome de host fictício `ip-xxx-xxx-xxx-xxx.ec2.internal`. O usuário está associado à unidade organizacional (OU) `admins` e com a senha `123456`:

```
ldapwhoami -x -w "123456" -D uid=presto,ou=admins,dc=ec2,dc=internal -H ldaps://ip-xxx-xxx-xxx-xxx.ec2.internal
```

Se a conta for válida e tiver as permissões adequadas, o comando retornará:

```
dn:uid=presto,ou=admins,dc=ec2,dc=internal
```

As configurações de exemplo em [Etapa 3: criar uma configuração JSON com as propriedades do Presto para LDAP](#) incluem essa conta para maior clareza, com exceção do exemplo de 5.10.0, em que ela não é compatível. Se o LDAP servidor usar associação anônima, remova os `internal-communication.ldap.user` pares `internal-communication.ldap.password` nome/valor e.

O nome LDAP distinto (DN) para usuários do Presto

Ao especificar a LDAP configuração do Presto, você especifica um padrão de vinculação que consiste `#{USER}` em uma unidade organizacional (OU) e componentes de domínio adicionais (DCs). O Presto `#{USER}` substitui pelo ID de usuário (UID) real de cada usuário durante a autenticação por senha para corresponder ao nome distinto (DN) que esse padrão de associação especifica. Você precisa do OUs que os usuários qualificados pertencem e delesDCs. Por exemplo, para permitir que os usuários da OU `admins` no domínio `corp.example.com` autenticarem para o Presto, especifique `#{USER},ou=admins,dc=corp,dc=example,dc=com` como o padrão de associação do usuário.

Note

Ao usar AWS CloudFormation, você precisa usar a função `Fn::Sub` para substituí-la pela ID `#{USER}` de usuário real (`()`). UID Para obter mais informações, consulte o tópico [Fn::Sub](#) no Guia do usuário do AWS CloudFormation .

Ao usar o Amazon EMR 5.10.0, você pode especificar somente um desses padrões. Usando o Amazon EMR 5.11.0 ou posterior, você pode especificar vários padrões separados por dois pontos (:). Os usuários que tentam autenticar para o Presto são comparados com o primeiro padrão, o segundo, e assim por diante. Para ver um exemplo, consulte [Etapa 3: criar uma configuração JSON com as propriedades do Presto para LDAP](#).

Etapa 2: definir uma configuração de segurança

Crie uma configuração de segurança com a criptografia em trânsito habilitada. Para obter mais informações, consulte [Criar uma configuração de segurança](#) no Amazon EMR Management Guide. Os artefatos de criptografia que você fornece ao configurar a criptografia em trânsito são usados

para criptografar a comunicação interna entre os nós Presto. Para obter mais informações, consulte [Fornecer certificados para criptografia de dados em trânsito](#). O certificado LDAP do servidor é usado para autenticar as conexões do cliente com o servidor Presto.

Etapa 3: criar uma configuração JSON com as propriedades do Presto para LDAP

Você usa a classificação `presto-config` de configuração para definir as propriedades do Presto para LDAP. O formato e o conteúdo do `presto-config` são ligeiramente diferentes dependendo da versão de EMR lançamento da Amazon e da instalação do Presto (PrestoDB ou Trino). Exemplos de diferenças de configuração serão fornecidos posteriormente nesta seção. Para obter mais informações, consulte [Configurar aplicações](#).

As etapas a seguir pressupõem que você salve os JSON dados em um arquivo, *MyPrestoConfig.json*. Se você usar o console, faça o upload do arquivo em um local seguro no Amazon S3 para que você possa referenciá-lo ao criar o cluster. Se você usar o AWS CLI, poderá referenciar o arquivo localmente.

Example Amazon EMR 6.1.0 e versões posteriores com Presto SQL (Trino)

O exemplo a seguir usa o nome do LDAP host de [Etapa 1: Reúna informações sobre seu LDAP servidor e copie o certificado do servidor para o Amazon S3](#) para se autenticar no LDAP servidor para vinculação. Dois padrões de associação de usuário são especificados, o que indica que os usuários dentro da `admins datascientists OU` e da `OU` no LDAP servidor estão qualificados para autenticação no servidor Trino como usuários. Os padrões de associação são separados com uma vírgula (:).

EMRAs versões 6.4.0 e posteriores da Amazon usam o novo nome Trino em vez de Presto. SQL Se você usa Trino, substitua *configuração prestosql* na seguinte classificação de configuração com `trino-config` e *prestosql-password-authenticator* com `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"
    }
  }
]

```

Example Amazon EMR 5.16.0 e versões posteriores

O exemplo a seguir usa o ID LDAP do usuário e a senha e o nome do LDAP host [Etapa 1: Reúna informações sobre seu LDAP servidor e copie o certificado do servidor para o Amazon S3](#) para autenticação no LDAP servidor para vinculação. Dois padrões de associação de usuário são especificados, o que indica que os usuários dentro da admins datascientists OU e da OU no LDAP servidor estão qualificados para autenticação no servidor Presto como usuários. Os padrões de associação são separados com uma vírgula (:).

```

[ {
  "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"
  }
}
]

```

Example Amazon EMR 5.11.0 a 5.15.0

O formato da classificação de configuração presto-config é um pouco diferente para essas versões. O exemplo a seguir especifica os mesmos parâmetros que o exemplo anterior.

```

[ {

```

```

    "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"
    }
  ]
}

```

Example Amazon EMR 5.10.0

O Amazon EMR 5.10.0 oferece suporte somente à vinculação anônima, portanto, essas entradas são omitidas. Além disso, apenas um único padrão de associação pode ser especificado.

```

[ {
  "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"
  }
}
]

```

Etapa 4: Crie o script para copiar o certificado LDAP do servidor e carregá-lo no Amazon S3

Crie um script que copie o arquivo do certificado em cada nó do cluster e o adicione ao repositório de chaves. Crie o script usando um editor de texto, salve-o e carregue-o no Amazon S3. Em [Etapa 5: criar um cluster](#), o arquivo de script é referenciado como `s3://MyBucket/LoadLDAPCert.sh`.

O script de exemplo a seguir usa a senha padrão do keystore, `changeit`. Recomendamos que você se conecte ao nó principal depois de criar o cluster e alterar a senha do keystore usando o comando `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
```

Etapa 5: criar um cluster

Ao criar o cluster, você especifica o Presto e outros aplicativos que deseja que EMR a Amazon instale. Os exemplos a seguir também fazem referência às propriedades de classificação de configuração em um JSON, mas você também pode especificar a classificação de configuração em linha.

Para criar um cluster Presto com LDAP autenticação usando o console da Amazon EMR

1. Navegue até o novo EMR console da Amazon e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte [Usar o console antigo](#).
2. Escolha Create cluster (Criar cluster), Go to advanced options (Ir para opções avançadas).
3. Escolha o Presto junto com outros aplicativos para EMR a Amazon instalar e, em Configuração de software, selecione a versão da Amazon EMR a ser usada. LDAPa autenticação é suportada somente com o Amazon EMR 5.10.0 e versões posteriores.
4. Em Editar configurações de software, escolha Carregar JSON do S3, insira a localização no Amazon S3 JSON do arquivo de configuração que você criou e, [Etapa 3: criar uma configuração JSON com as propriedades do Presto para LDAP](#) em seguida, escolha Avançar.
5. Configure os recursos de hardware e rede do cluster e, em seguida, selecione Next (Próximo).
6. Selecione Bootstrap Actions (Ações de bootstrap). Em Add bootstrap action (Adicionar ação de bootstrap), selecione Custom action (Personalizar ação) e, em seguida, selecione Configure and add (Configurar e adicionar).
7. Insira um nome para a ação de bootstrap, insira a localização do script que você criou [Etapa 4: Crie o script para copiar o certificado LDAP do servidor e carregá-lo no Amazon S3](#), por exemplo, `s3://MyBucket/LoadLDAPCert.sh` e escolha Adicionar.
8. Em General Options (Opções gerais), Tags e Additional Options (Opções adicionais), selecione as configurações adequadas para o seu aplicativo e, em seguida, selecione Next (Próximo).
9. Selecione Authentication and encryption (Autenticação e criptografia) e, em seguida, selecione a Security configuration (Configuração de segurança) que você criou em [Etapa 2: definir uma configuração de segurança](#).

10. Selecione outras opções de segurança conforme adequado para o seu aplicativo e, em seguida, selecione Create cluster (Criar cluster).

Para criar um cluster Presto com LDAP autenticação usando o AWS CLI

- Use o comando `aws emr create-cluster`. No mínimo, especifique o aplicativo do Presto e também a classificação de configuração, o script de bootstrap e a configuração de segurança do Presto que você criou nas etapas anteriores. O exemplo a seguir faz referência ao arquivo de configuração como um JSON arquivo salvo no mesmo diretório em que você executa o comando. Por outro lado, o script de bootstrap deve ser salvo no Amazon S3. O exemplo a seguir usa `s3://MyBucket/LoadLDAPCert.sh`.

Note

Os caracteres de continuação de linha do Linux (`\`) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (`^`).

```
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
```

Ativar o modo estrito do Presto

Em determinadas situações, consultas de longa duração podem gerar altos custos e fazer com que EMR a Amazon use mais recursos de cluster. Isso retira recursos de outras workloads do cluster. Com EMR as versões 6.8 e posteriores da Amazon, você pode usar um recurso de modo estrito que rejeita ou avisa sobre os seguintes tipos de consultas de longa duração:

- Consultas sem predicados nas colunas particionadas que resultam em verificações de tabelas de grandes quantidades de dados

- Consultas com junções cruzadas entre duas tabelas grandes
- Consultas que classificam um grande número de linhas sem limite

Depois que o Presto otimiza por completo o plano de consulta, o modo estrito é executado. Para usar e personalizar o modo estrito de acordo com suas necessidades de consulta, configure o Presto das maneiras a seguir.

Configurações do Presto para modo estrito

Configuração	Descrição	Padrão
<code>strict-mode-enabled</code>	Ativa e desativa o modo estrito. Um valor de <code>true</code> indica que o modo estrito está ativado.	<code>false</code>
<code>strict-mode-fail-query</code>	Rejeita consultas se o modo estrito detecta prováveis consultas de execução prolongada. Se <code>false</code> , a Amazon EMR apenas emite um aviso.	<code>false</code>
<code>strict-mode-restrictions</code>	Especifica as restrições a serem aplicadas quando o modo estrito está ativado. O modo estrito suporta as seguintes restrições: MANDATORY PARTITION _ DISALLOW _ PREDICATE JOIN, CROSS _ _ e LIMITED _ SORT.	MANDATORY_PARTITION_PREDICATE, DISALLOW_CROSS_JOIN, LIMITED_SORT

Para testar o modo estrito, você pode substituir essas configurações ou defini-las como propriedades da sessão ao usar o cliente do Presto.

Para definir a configuração na criação do cluster com o AWS Management Console

1. Escolha Create cluster e selecione Amazon EMR versão 6.8.0 e Presto ou Trino. Para obter mais informações, consulte [Instalar PrestoDB e Trino](#).
2. Especifique diretamente as propriedades de configuração para o modo estrito ou faça o upload de um JSON arquivo para o Amazon S3. Como alternativa, selecione o para seu metastore. Especifique suas sub-redesVPC, ações de bootstrap, key pair e grupo de segurança. Para criar o cluster, escolha Criar cluster.
3. Faça login no nó primário do cluster e execute `presto-cli` ou `trino-cli`.
4. Envie suas consultas. O modo estrito valida cada consulta e determina se ela é de execução prolongada. Dependendo da sua `strict-mode-fail-query` configuração, a Amazon EMR rejeita a consulta ou emite um aviso.
5. Ao concluir as consultas, encerre o cluster e exclua seus recursos.

Para definir a configuração em um cluster em execução com o AWS CLI

1. Faça login no nó primário do seu cluster com o AWS CLI e execute `presto-cli` ou `trino-cli`.
2. Execute os comandos a seguir com os valores desejados.

```
set session strict_mode_enabled = true;
set session strict_mode_fail_query = false;
set session strict_mode_restrictions = 'DISALLOW_CROSS_JOIN,LIMITED_SORT';
```

Considerações

Quando usar o modo estrito, considere os seguintes pontos:

- Em alguns casos, o modo estrito pode rejeitar consultas de execução curta que não consomem muitos recursos. Por exemplo, consultas em tabelas pequenas não aplicam filtragem dinâmica nem substituem junções internas por junções cruzadas. Isso pode fazer com que a consulta use o predicado de partição obrigatória ou proíba junção cruzada. Quando isso acontece, o modo estrito rejeita a consulta.
- A verificação do modo estrito é aplicada somente nos tipos `SELECTINSERT`, `SELECT`, `CREATE TABLE AS` e de `EXPLAIN ANALYZE` consulta.
- Você só pode usar a restrição no predicado de partição obrigatória com o conector do Hive.

Lidar com a perda de instâncias spot no Presto

Com as Instâncias Spot na AmazonEMR, você pode executar cargas de trabalho de big data na EC2 capacidade disponível da Amazon a um custo reduzido. Em troca do menor custo, a Amazon EC2 pode interromper as Instâncias Spot com uma notificação de dois minutos. Quando você encerra um nó, o Presto pode levar até 10 minutos para retornar um erro. Isso causa atrasos desnecessários nos relatórios de erros e possíveis novas tentativas. O encerramento rápido é um atributo que permite o controle da forma como o Presto manipula nós encerrados.

O trabalho do coordenador do Presto é acompanhar todos os nós de processamento com sondagens regulares dos seus status. Sem uma rescisão rápida, o coordenador não consulta o YARN NodeManager para saber o status de cada nó. Isso pode resultar em um longo loop de novas tentativas antes que a consulta apresente falha. Com o encerramento rápido, o coordenador do Presto consulta o status do nó NodeManager assim que a pesquisa não chega ao host. Se NodeManager mostrar que o nó está inativo, o Presto abandona novas tentativas, falha na consulta e retorna um erro. `NODE_DECOMMISSIONED`

O conjunto de parâmetros de configuração a seguir permite o controle e a personalização do comportamento do Presto no caso do encerramento do nó.

Configurações do Presto para tratamento de falhas em nós

Configuração	Descrição	Padrão
<code>query.remote-task.max-backoff-duration</code>	Período durante o qual o coordenador continua tentando obter o status da tarefa remota nos nós de processamento.	10 minutos
<code>query.remote-task.quick-terminate-on-connect-exception</code>	Ativa a falha rápida do nó se o coordenador não consegue acessar o nó ou não consegue se conectar ao worker executado nesse nó. O valor de <code>query.remote-task.terminate-on-connect-exception</code> determina se o coordenador	true

Configuração	Descrição	Padrão
	<p>or deve acessar o nó ou se conectar ao worker.</p> <p>O nó falha na consulta e a Amazon EMR remove o nó da lista de trabalhadores disponíveis. Quando isso acontece, você não pode usar o nó para programar novas consultas.</p> <p>Quando você define esse valor como <code>false</code>, o Presto volta ao comportamento anterior, em que o coordenador do Presto novamente tenta acessar o nó (para <code>query.remote-task.max-backoff-duration</code>) antes de marcar o nó como indisponível e indica falha na consulta em andamento no nó.</p>	
<code>query.remote-task.terminate-on-connect-exception</code>	<p>Especifica se a Amazon EMR deve ser um nó se o host estiver acessível, mas o coordenador não conseguir se conectar ao processo de trabalho do host. Ao definir esse valor como <code>true</code>, você ativa falha na consulta rápida se o host está inacessível.</p>	<code>false</code>

Execução tolerante a falhas no Trino

A execução tolerante a falhas é um mecanismo no Trino que um cluster pode usar para mitigar falhas na consulta. Para fazer isso, ele repete as consultas ou as tarefas dos seus componentes quando elas apresentam falha. Quando a execução tolerante a falhas é ativada, os dados intermediários de troca são armazenados em spool e outro worker pode reutilizá-los no caso de uma interrupção do worker ou outra falha durante a execução da consulta.

Para obter mais informações sobre execução tolerante a falhas no Trino, consulte [Project Tardigrade entrega na velocidade do Trino para os primeiros ETL usuários no blog do Trino](#).

Configuração

A execução tolerante a falhas está desativada por padrão. Para ativar o atributo, defina a propriedade de configuração `retry-policy` na classificação `trino-config` como `QUERY` ou `TASK` com base na política de novas tentativas desejada, da seguinte maneira.

```
{"classification":  
  "trino-config",  
  "properties":  
    {  
      "retry-policy":  
        "QUERY"  
    }  
}
```

Uma política de novas tentativas **QUERY** instrui o Trino a repetir uma consulta automaticamente quando ocorre um erro em um nó de processamento. Recomendamos que você use uma política de novas tentativas `QUERY` quando a maior parte da workload do cluster do Trino incluir muitas consultas pequenas.

Uma política de novas tentativas **TASK** instrui o Trino a tentar novamente tarefas de consulta individuais em caso de falha. Recomendamos essa política quando o Trino executa grandes consultas em lotes. O cluster pode tentar novamente tarefas menores na consulta com mais eficiência em vez de tentar novamente a consulta inteira.

Gerenciador de trocas

Um gerenciador de trocas armazena e gerencia dados em spool para execuções tolerantes a falhas. Ele usa armazenamento externo para armazenar dados despejados além do tamanho do buffer na

memória. Você pode configurar um gerenciador do Exchange baseado em sistema de arquivos que armazena dados em pool em um local específico, como Amazon S3, sistemas compatíveis com Amazon S3 ou. HDFS

As EMR versões 6.9.0 e posteriores da Amazon incluem a `trino-exchange-manager` classificação para configurar o gerenciador de câmbio. Essas versões também oferecem suporte HDFS para spooling.

Configurar o gerenciador de trocas

Use a classificação de configuração `trino-exchange-manager` para configurar um gerenciador de trocas. Essa classificação cria internamente um arquivo de configuração `etc/exchange-manager.properties` no coordenador e em todos os nós de processamento. A classificação também define a propriedade de configuração `exchange-manager.name` como `filesystem`.

Por padrão, a Amazon EMR lança a versão 6.9.0 e, posteriormente, é usada HDFS como gerente do Exchange. HDFS está disponível nos EMR EC2 clusters da Amazon, e o spooling ocorre no `trino-exchange/` diretório por padrão. Para usar as configurações padrão, faça a seguinte configuração:

```
{"Classification":  
  "trino-exchange-manager"  
}
```

Se você quiser fornecer um local personalizado, defina as seguintes propriedades na classificação `trino-exchange-manager`:

- Defina `exchange.use-local-hdfs` como `true`.
- `exchange.base-directories` Defina como a localização do diretório personalizado em HDFS, por exemplo, `exchange.base-directories=/exchange`. Se o diretório personalizado ainda não estiver dentro HDFS, a Amazon EMR o criará.

HDFS configurações do Exchange Manager

Com base nos resultados dos testes internos, recomendamos que você passe para o local HDFS para melhorar o desempenho da consulta em comparação com outros sistemas de arquivos baseados em nuvem. Você pode definir as seguintes configurações para o gerenciador de câmbio com HDFS.

Configuração	Descrição	Configuração padrão
<code>exchange.hdfs.block-size</code>	Tamanho do bloco para HDFS armazenamento	4 MB
<code>hdfs.config.resources</code>	Lista de caminhos de arquivo a serem configurados HDFS	Se <code>exchange.use-local-hdfs</code> é <code>true</code> , são usados os caminhos para arquivos <code>core-site.xml</code> , <code>hdfs-site.xml</code> ; caso contrário, <code>null</code>

Para obter propriedades adicionais de configuração de execução tolerantes a falhas e para obter informações sobre como configurar o Amazon S3 ou outros sistemas compatíveis com o Amazon S3 para spooling, consulte a página [Execução tolerante a falhas](#) da documentação do Trino.

Considerações e limitações

- Se você habilitar a execução tolerante a falhas, ela desativará as operações de `write` para conectores que não são compatíveis com `write` quando `retry-policy` está definida. A partir da EMR versão 6.9.0 da Amazon, os conectores Delta Lake, Hive e Iceberg dão suporte às operações com `write retry-policy`
- Se você usa o gerenciador de trocas e executa operações de E/S caras, suas consultas podem ter diminuição da performance enquanto o gerenciador de trocas transfere os dados intermediários para um armazenamento externo.

Usar a escalabilidade automática do Presto com desativação tranquila

As EMR versões 5.30.0 e posteriores da Amazon incluem um recurso que você pode usar para definir um período de carência para determinadas ações de escalabilidade. O período de carência possibilita que as tarefas do Presto continuem em execução antes que o nó seja encerrado devido a uma ação de redimensionamento para reduzir a escala horizontalmente ou a uma solicitação de política de escalabilidade automática. Para obter mais informações sobre regras de escalabilidade, consulte [Entendendo as regras de escalabilidade automática](#) no Amazon EMR Management Guide.

A escalabilidade automática do Presto com desativação tranquila impede que novas tarefas sejam programadas em um nó que está sendo desativado e, ao mesmo tempo, permite que as tarefas que já estão sendo executadas sejam concluídas antes que o tempo limite de desligamento seja atingido. As consultas em execução concluirão a execução antes que o nó seja desativado. A escalabilidade automática não é compatível com frotas de instâncias.

Você pode controlar quanto tempo deve ser permitido até a conclusão das tarefas do Presto após o recebimento de uma solicitação de desligamento de escalabilidade automática. Por padrão, o tempo limite de desligamento da Amazon EMR é de 0 minutos, o que significa que a Amazon encerra EMR imediatamente o nó e todas as tarefas do Presto em execução nele, se exigido por uma solicitação de escalabilidade. Para definir um tempo limite maior para as tarefas do Presto na Amazon, EMR a fim de permitir que as consultas sejam concluídas antes de reduzir a escala de um cluster, use a classificação de `presto-config` configuração para definir o `graceful-shutdown-timeout` parâmetro com um valor em segundos ou minutos maior que zero. Para obter mais informações, consulte [Configurar aplicações](#).

Por exemplo, aumentar o valor de `graceful-shutdown-timeout` para `"30m"` especifica um tempo limite de 30 minutos. Após o término do tempo limite de desligamento, o nó marcado para desativação será encerrado de maneira forçada se estiver aguardando a conclusão das tarefas de consulta e a consulta apresentará falha. Se as tarefas de consulta terminarem em cinco minutos, o nó marcado para descomissionamento será encerrado em cinco minutos, desde que outros YARN aplicativos tenham concluído a execução.

Example Exemplo de configuração de escalabilidade automática do Presto com desativação tranquila

Substitua o valor `graceful-shutdown-timeout` pelo número de minutos apropriado para a configuração. Não há um valor máximo. O exemplo abaixo define um valor de tempo limite de 1800 segundos (30 minutos).

```
[
  {
    "classification": "presto-config",
    "properties": {
      "graceful-shutdown-timeout": "1800s"
    }
  }
]
```

Limitações

O PrestoDB Graceful Decommission não funciona EMR em clusters em que a conectividade está desativada, como HTTP quando está definido como `http-server.http.enabled false`. O Trino não é compatível com a desativação tranquila, independentemente da configuração de `http-server.http.enabled`.

Considerações sobre o Presto na Amazon EMR

Considere as seguintes limitações ao executar o [Presto](#) na AmazonEMR.

Executável da linha de comando do Presto

Na AmazonEMR, o PrestoDB e o Trino usam o mesmo executável de linha de comando `presto-cli`, como no exemplo a seguir.

```
presto-cli --catalog hive
```

Propriedades de implantação do Presto não configuráveis

A versão da Amazon EMR que você usa determina as configurações de implantação do Presto que estão disponíveis. Para obter mais informações sobre essas propriedades de configuração, consulte [Implantar o Presto](#) na documentação do Presto. A tabela a seguir mostra as diferentes opções de configuração para arquivos de `properties` do Presto.

Arquivo	Configurável
<code>log.properties</code>	<p>PrestoDB: configurável nas versões 4.0.0 e posteriores da EMR Amazon. Use a classificação para a configuração <code>presto-log</code> .</p> <p>Trino (PrestoSQL): configurável nas EMR versões 6.1.0 e posteriores da Amazon. Use a classificação de configuração <code>prestoql-log</code> ou <code>trino-log</code> .</p>
<code>config.properties</code>	<p>PrestoDB: configurável nas versões 4.0.0 e posteriores da EMR Amazon. Use a classificação para a configuração <code>presto-config</code> .</p>

Arquivo	Configurável
	Trino (PrestoSQL): configurável nas EMR versões 6.1.0 e posteriores da Amazon. Use a classificação de configuração <code>prestoql-config</code> ou <code>trino-config</code> .
<code>hive.properties</code>	<p>PrestoDB: configurável nas versões 4.1.0 e posteriores da EMR Amazon. Use a classificação para a configuração <code>presto-connector-hive</code> .</p> <p>Trino (PrestoSQL): configurável nas EMR versões 6.1.0 e posteriores da Amazon. Use a classificação de configuração <code>prestoql-connector-hive</code> ou <code>trino-connector-hive</code> .</p>
<code>node.properties</code>	<p>PrestoDB: configurável na EMR Amazon versão 5.6.0 e posterior. Use a classificação para a configuração <code>presto-node</code> .</p> <p>Trino (PrestoSQL): configurável nas EMR versões 6.1.0 e posteriores da Amazon. Use a classificação de configuração <code>prestoql-node</code> ou <code>trino-node</code> .</p>
<code>jvm.config</code>	Não configurável.

Instalação do PrestoDB e do Trino

O nome da aplicação, Presto, continua a ser usado para instalar o PrestoDB em clusters. Para instalar o Trino em clusters, use o nome do aplicativo Trino (ou Presto SQL em versões mais antigas da Amazon). EMR

Você pode instalar o PrestoDB ou o Trino, mas não pode instalar os dois em um único cluster. Se o PrestoDB e o Trino forem especificados na tentativa de criação de um cluster, ocorrerá um erro de validação e a solicitação de criação do cluster falhará.

EMRFS e configuração do PrestoOS3 FileSystem

Com EMR as versões 5.12.0 e posteriores da Amazon, o PrestoDB pode usar EMRFS. Essa é a configuração padrão. EMRFS também é o sistema de arquivos padrão para Trino (PrestoSQL) EMR nas versões 6.1.0 e posteriores da Amazon. Para obter mais informações, consulte [EMR File System \(EMRFS\)](#) no Amazon EMR Management Guide. Nas versões anteriores da Amazon EMR, o PrestoOS3 FileSystem é a única opção de configuração.

Você pode usar uma configuração de segurança para configurar a criptografia de EMRFS dados no Amazon S3. Você também pode usar IAM funções para EMRFS solicitações ao Amazon S3. Para obter mais informações, consulte [Entendendo as opções de criptografia](#) e [Configurando IAM funções para EMRFS solicitações ao Amazon S3](#) no Amazon EMR Management Guide.

Note

Se você consultar dados subjacentes no Amazon S3 com a EMR versão 5.12.0 da Amazon, podem ocorrer erros do Presto. Isso acontece porque o Presto não consegue obter valores de classificação de configuração em `emrfs-site.xml`. Como solução alternativa, crie um subdiretório `emrfs` em `usr/lib/presto/plugin/hive-hadoop2/` e crie um link simbólico em `usr/lib/presto/plugin/hive-hadoop2/emrfs` para o arquivo `/usr/share/aws/emr/emrfs/conf/emrfs-site.xml` existente. Em seguida, reinicie o processo `presto-server` (`sudo presto-server stop` seguido por `sudo presto-server start`).

Você pode substituir o EMRFS padrão e usar o FileSystem PrestoOS3 em vez disso. Para fazer isso, use a classificação de configuração `presto-connector-hive` para definir `hive.s3-file-system-type` como `PRESTO`, conforme mostrado no exemplo a seguir. Para obter mais informações, consulte [Configurar aplicações](#).

```
[
  {
    "Classification": "presto-connector-hive",
    "Properties": {
      "hive.s3-file-system-type": "PRESTO"
    }
  }
]
```

Se você usa o PrestoOS3FileSystem, use a classificação de `presto-connector-hive` configuração ou o Trino configure `trino-connector-hive` as propriedades do PrestoOS3.FileSystem Para obter mais informações sobre as propriedades disponíveis, consulte [Configuração do Amazon S3](#) na seção Conector do Hive da documentação do Presto. Essas configurações não se aplicam EMRFS a.

Configuração padrão para personificação do usuário final

Por padrão, EMR as versões 5.12.0 e posteriores da Amazon permitem a representação do usuário final para acesso a HDFS Para obter mais informações, consulte [Personificação de usuário final](#) na documentação do Presto. Para alterar essa configuração usando a classificação de configuração `presto-config`, defina a propriedade `hive.hdfs.impersonation.enabled` como `false`.

Porta padrão para interface Web do Presto

Por padrão, a Amazon EMR configura a interface web do Presto no coordenador do Presto para usar a porta 8889 (para PrestoDB e Trino). Para alterar a porta, use a classificação de configuração `presto-config` para definir a propriedade `http-server.http.port`. Para obter mais informações, consulte [Propriedades de configuração](#) na seção Implantar o Presto na documentação do Presto.

Problema com a execução do bucket do Hive em algumas versões

O Presto versão 152.3 tem um problema com a execução do bucket do Hive que prejudica significativamente o desempenho das consultas do Presto em algumas circunstâncias. EMRAs versões 5.0.3, 5.1.0 e 5.2.0 da Amazon incluem essa versão do Presto. Para reduzir o problema, use a classificação de configuração `presto-connector-hive` para definir a propriedade `hive.bucket-execution` como `false`, conforme mostrado no exemplo a seguir.

```
[
  {
    "Classification": "presto-connector-hive",
    "Properties": {
      "hive.bucket-execution": "false"
    }
  }
]
```

Histórico de versões do Presto

A tabela a seguir lista a versão do Presto incluída em cada versão de lançamento da AmazonEMR, junto com os componentes instalados com o aplicativo. Para obter as versões dos componentes em cada versão, consulte a seção Versão de componentes da sua versão em [Versões de lançamento do Amazon EMR 7.x](#), [Versões de lançamento do Amazon EMR 6.x](#) ou [Versões de lançamento do Amazon EMR 5.x](#).

Informações de versão do Presto

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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
emr-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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

Gravadora Amazon EMR Release	Versão do Presto	Componentes instalados com o 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

A tabela a seguir lista a versão do Trino (PrestoSQL) incluída em cada versão de lançamento da AmazonEMR, junto com os componentes instalados com o aplicativo. O Presto SQL mudou seu nome para Trino a partir da versão 351.

Informações sobre a versão do Trino (PrestoSQL)

Gravadora Amazon EMR Release	Versão Trino (PrestoSQL)	Componentes instalados com Trino (SQLPresto)
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-nodema

Gravadora Amazon EMR Release	Versão Trino (PrestoSQL)	Componentes instalados com Trino (SQLPresto)
		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

Gravadora Amazon EMR Release	Versão Trino (PrestoSQL)	Componentes instalados com Trino (SQLPresto)
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

Gravadora Amazon EMR Release	Versão Trino (PrestoSQL)	Componentes instalados com Trino (SQLPresto)
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

Gravadora Amazon EMR Release	Versão Trino (PrestoSQL)	Componentes instalados com Trino (SQLPresto)
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

Gravadora Amazon EMR Release	Versão Trino (PrestoSQL)	Componentes instalados com Trino (SQLPresto)
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

Gravadora Amazon EMR Release	Versão Trino (PrestoSQL)	Componentes instalados com Trino (SQLPresto)
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

Gravadora Amazon EMR Release	Versão Trino (PrestoSQL)	Componentes instalados com Trino (SQLPresto)
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

Gravadora Amazon EMR Release	Versão Trino (PrestoSQL)	Componentes instalados com Trino (SQLPresto)
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

Gravadora Amazon EMR Release	Versão Trino (PrestoSQL)	Componentes instalados com Trino (SQLPresto)
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

Gravadora Amazon EMR Release	Versão Trino (PrestoSQL)	Componentes instalados com Trino (SQLPresto)
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

Gravadora Amazon EMR Release	Versão Trino (PrestoSQL)	Componentes instalados com Trino (SQLPresto)
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

Notas de lançamento do Trino (PrestoSQL) por versão

- [Amazon EMR 6.9.0 - Notas de lançamento do Trino \(SQLPresto\)](#)

Amazon EMR 6.9.0 - Notas de lançamento do Trino (SQLPresto)

Amazon EMR 6.9.0 - Novos recursos do Trino (SQLPresto)

- Para dar suporte a consultas de longa execução, o Trino agora inclui um mecanismo de execução tolerante a falhas. A execução tolerante a falhas atenua as falhas nas consultas ao tentar novamente as consultas com falha ou as tarefas dos seus componentes. Para obter mais informações, consulte [Execução tolerante a falhas no Trino](#).

Amazon EMR 6.9.0 - Mudanças em Trino (Presto) SQL

Amazon EMR 6.9.0 - Mudanças no PrestoDB

Tipo	Descrição
Upgrade	Atualização do PrestoDB para 0.276
Upgrade	Suporte para Hadoop 3.3.3
Upgrade	Atualização do Hudi para 0.12.1
Atributo	Integração entre Amazon EMR e Presto AWS Lake Formation para uso de cargas de trabalho interativas. GCSC API
Atributo	Adicionadas configurações relacionadas ao Kerberos nas Configurações de segurança do PrestoDB para habilitar o Keberos.
Correção de bugs	Reverteu o OSS pull request #18115, que foi adicionado para reduzir o número de <code>hdfsConfiguration</code> cópias. Isso fez com que a HDFS configuração fosse copiada incorretamente ao usar EMRFS nossas tabelas Hudi.

Amazon EMR 6.9.0 - Mudanças no Trino

Tipo	Descrição
Upgrade	Atualização do Trino para 398
Upgrade	Suporte para Hadoop 3.3.3
Atributo	Suporte Tardigrade: adicione suporte para spooling do Exchange no Amazon HDFS S3. Para ter mais informações, consulte Execução tolerante a falhas no Trino
Correção de bugs	Quando o Iceberg do Trino é usado e o catálogo do Glue está habilitado, evite adicionar o uri do metastore em <code>iceberg.properties</code> .

Amazon EMR 6.9.0 - Problemas conhecidos do Trino (SQLPresto)

- Para a EMR versão 6.9.0 da Amazon, o Trino não funciona em clusters habilitados para o Apache Ranger. Se você precisar usar o Trino com o Ranger, entre em contato com o [AWS Support](#).

Apache Spark

O [Apache Spark](#) é uma estrutura de processamento distribuído e um modelo de programação que ajuda você a fazer aprendizado de máquina, processamento de streams ou análise gráfica com clusters da AmazonEMR. Semelhante ao Apache Hadoop, o Spark é um sistema de processamento distribuído de código-fonte aberto comumente utilizado para cargas de trabalho Big Data. No entanto, o Spark tem várias diferenças notáveis em relação ao MapReduce Hadoop. O Spark tem um mecanismo otimizado de execução de grafos acíclicos direcionados (DAG) e armazena ativamente os dados na memória, o que pode aumentar o desempenho, especialmente para determinados algoritmos e consultas interativas.

O Spark oferece suporte nativo a aplicativos escritos em Scala, Python e Java. [Ele também inclui várias bibliotecas totalmente integradas para SQL \(Spark SQL\), aprendizado de máquina \(MLlib\), processamento de fluxo \(streaming Spark\) e processamento gráfico \(GraphX\)](#). Essas ferramentas facilitam o aproveitamento da estrutura do Spark para uma grande variedade de casos de uso.

Você pode instalar o Spark em um EMR cluster da Amazon junto com outros aplicativos do Hadoop, e ele também pode aproveitar o sistema de EMR arquivos da Amazon (EMRFS) para acessar diretamente os dados no Amazon S3. O Hive também é integrado ao Spark para que você possa usar um HiveContext objeto para executar scripts do Hive usando o Spark. Um contexto do Hive está incluído no spark-shell como `sqlContext`.

Para ver um exemplo de tutorial sobre como configurar um EMR cluster com o Spark e analisar um conjunto de dados de amostra, consulte [Tutorial: Getting started with Amazon EMR](#) no blog AWS News.

Important

[A versão 2.3.1 do Apache Spark, disponível a partir da EMR versão 5.16.0 da Amazon, endereça -2018-8024 e -2018-1334. CVE CVE](#) Recomendamos que você migre as versões anteriores do Spark para a versão 2.3.1 ou posteriores.

A tabela a seguir lista a versão do Spark incluída na versão mais recente da série Amazon EMR 7.x, junto com os componentes que a Amazon EMR instala com o Spark.

Para a versão dos componentes instalados com o Spark nesta versão, consulte [Versões de componentes da versão 7.2.0](#).

Informações sobre a versão do Spark para o emr-7.2.0

Gravadora EMR de lançamento da Amazon	Versão do Spark	Componentes instalados com o 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-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


A tabela a seguir lista a versão do Spark incluída na versão mais recente da série EMR 6.x da Amazon, junto com os componentes que a Amazon EMR instala com o Spark.

Para obter a versão dos componentes instalados com o Spark nessa versão, consulte [Release 6.15.0 Component Versions](#).

Informações de versão do Spark para o emr-6.15.0

Gravadora EMR de lançamento da Amazon	Versão do Spark	Componentes instalados com o 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-httfs-server, hadoop-

Gravadora EMR de lançamento da Amazon	Versão do Spark	Componentes instalados com o 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

A EMR versão 6.8.0 da Amazon vem com o Apache Spark 3.3.0. Esta versão do Spark usa o Apache Log4j 2 e o arquivo `log4j2.properties` para configurar o Log4j nos processos do Spark. Se você usa o Spark no cluster ou cria EMR clusters com parâmetros de configuração personalizados e deseja atualizar para a EMR versão 6.8.0 da Amazon, você deve migrar para a nova classificação de `spark-log4j2` configuração e formato de chave do Apache Log4j 2. Para obter mais informações, consulte [Migrar do Apache Log4j 1.x para Log4j 2.x](#).

A tabela a seguir lista a versão do Spark incluída na versão mais recente da série Amazon EMR 5.x, junto com os componentes que a Amazon EMR instala com o Spark.

Para a versão dos componentes instalados com o Spark nesta versão, consulte Versões de componentes da [versão 5.36.2](#).

Informações sobre a versão do Spark para o emr-5.36.2

Gravadora EMR de lançamento da Amazon	Versão do Spark	Componentes instalados com o 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, hadoop-hdfs-library, hadoop-

Gravadora EMR de lançamento da Amazon	Versão do Spark	Componentes instalados com o Spark
		hdfs-namenode, hadoop- httpfs-server, hadoop-kms- server, hadoop-yarn-nodemanager, hadoop-yarn-resourceman- ager, hadoop-yarn-timeline-server, hudi, hudi- spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn- slave

Tópicos

- [Criação de um cluster com o Apache Spark](#)
- [Execute aplicativos Spark com o Docker na Amazon 6.x EMR](#)
- [Use o AWS Glue Data Catalog como metastore para o Spark SQL](#)
- [Configurar o Spark](#)
- [Otimizar a performance do Spark](#)
- [Armazenamento em Cache de Fragmentos Resultantes do Spark](#)
- [Use o Nvidia RAPIDS Accelerator para Apache Spark](#)
- [Acessar o shell do Spark](#)
- [Use o Amazon SageMaker Spark para aprendizado de máquina](#)
- [Escrever uma aplicação do Spark](#)
- [Melhorar a performance do Spark com Amazon S3](#)
- [Adicionar uma etapa do Spark](#)
- [Visualizar o histórico de aplicações do Spark](#)
- [Acesse a web do Spark UIs](#)
- [Usando o conector Amazon Kinesis Data Streams de streaming estruturado do Spark](#)
- [Usando a integração do Amazon Redshift para o Apache Spark com a Amazon EMR](#)
- [Histórico de versões do Spark](#)

Criação de um cluster com o Apache Spark

O procedimento a seguir cria um cluster com o [Spark](#) instalado usando as Opções rápidas no EMR console da Amazon.

Como alternativa, use as Opções avançadas para personalizar ainda mais sua configuração de cluster ou para enviar etapas para instalar aplicações de forma programática e depois executar aplicações personalizadas. Com qualquer opção de criação de cluster, você pode escolher usar o AWS Glue como sua SQL metastore do Spark. Consulte [Use o AWS Glue Data Catalog como metastore para o Spark SQL](#) Para mais informações.

Para iniciar um cluster com o Spark instalado

1. Abra o EMR console da Amazon em <https://console.aws.amazon.com/emr>.
2. Escolha Criar cluster para usar Opções rápidas.
3. Insira um Nome de cluster. O nome do cluster não pode conter os caracteres <, >, \$, | ou `(crase).
4. Em Configuração de software, escolha uma opção de Versão.
5. Em Aplicações, escolha o pacote de aplicações do Spark.
6. Selecione outras opções conforme necessário e escolha Create cluster (Criar cluster).

Note

Para configurar o Spark quando estiver criando o cluster, consulte [Configurar o Spark](#).

Para iniciar um cluster com o Spark instalado usando o AWS CLI

- Crie o cluster usando o comando a seguir.

```
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

Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

Para iniciar um cluster com o Spark instalado usando o SDK for Java

Especifique o Spark como um aplicativo com `SupportedProductConfig` usado em `RunJobFlowRequest`.

- O seguinte exemplo mostra como criar um cluster com o Spark usando 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());
    }
}

```

Execute aplicativos Spark com o Docker na Amazon 6.x EMR

Com o Amazon EMR 6.0.0, os aplicativos Spark podem usar contêineres Docker para definir suas dependências de biblioteca, em vez de instalar dependências nas instâncias individuais da Amazon EC2 no cluster. Para executar o Spark com o Docker, primeiro é necessário configurar o

registro do Docker e definir parâmetros adicionais ao enviar um aplicativo do Spark. Para obter mais informações, consulte [Configurar a integração do Docker](#).

Quando o aplicativo é enviado, YARN invoca o Docker para extrair a imagem do Docker especificada e executar o aplicativo Spark dentro de um contêiner do Docker. Isso permite definir e isolar dependências facilmente. Isso reduz o tempo de inicialização ou preparação de instâncias no EMR cluster da Amazon com as bibliotecas necessárias para a execução do trabalho.

Considerações ao executar o Spark com o Docker

Ao executar o Spark com o Docker, verifique se os seguintes pré-requisitos são atendidos:

- O `docker` pacote e CLI são instalados somente nos nós principais e de tarefas.
- No Amazon EMR 6.1.0 e versões posteriores, você também pode instalar o Docker em um nó primário usando os seguintes comandos.
 - ```
sudo yum install -y docker
sudo systemctl start docker
```
- O `spark-submit` comando sempre deve ser executado a partir de uma instância primária no EMR cluster da Amazon.
- Os registros do Docker usados para resolver imagens do Docker devem ser definidos usando a Classificação API com a chave de `container-executor` classificação para definir parâmetros adicionais ao iniciar o cluster:
  - `docker.trusted.registries`
  - `docker.privileged-containers.registries`
- Para executar um aplicativo do Spark em um contêiner do Docker, as seguintes opções de configuração são necessárias:
  - `YARN_CONTAINER_RUNTIME_TYPE=docker`
  - `YARN_CONTAINER_RUNTIME_DOCKER_IMAGE={DOCKER_IMAGE_NAME}`
- Ao usar ECR a Amazon para recuperar imagens do Docker, você deve configurar o cluster para se autenticar. Para fazer isso, é necessário usar a seguinte opção de configuração:
  - `YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG={DOCKER_CLIENT_CONFIG_PATH_HDFS_EM}`
- No Amazon EMR 6.1.0 e versões posteriores, você não precisa usar o comando listado `YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG={DOCKER_CLIENT_CONFIG_PATH_ON_HDFS}` quando o recurso de autenticação ECR automática está ativado.

- Qualquer imagem do Docker usada com o Spark deve ter o Java instalado na imagem do Docker.

Para obter mais informações sobre os pré-requisitos, consulte [Configurar a integração do Docker](#).

## Criar uma imagem do Docker

As imagens do Docker são criadas usando um arquivo do Docker, que define os pacotes e a configuração a serem incluídos na imagem. Os dois exemplos a seguir são usados por Dockerfiles e PySpark SparkR.

### PySpark Arquivo Docker

As imagens do Docker criadas a partir desse Dockerfile incluem o Python 3 e o pacote Python. NumPy Esse Dockerfile usa o Amazon Linux 2 e o 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"
```

### Arquivo do Docker do SparkR

As imagens do Docker criadas a partir desse Dockerfile incluem R e o pacote. randomForest CRAN  
Esse Dockerfile inclui o Amazon Linux 2 e o 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')"
```

Para obter mais informações sobre a sintaxe do arquivo do Docker, consulte a [documentação de referência do arquivo do Docker](#).

## Usando imagens Docker da Amazon ECR

O Amazon Elastic Container Registry (AmazonECR) é um registro de contêineres do Docker totalmente gerenciado, que facilita o armazenamento, o gerenciamento e a implantação de imagens de contêineres do Docker. Ao usar a AmazonECR, o cluster deve ser configurado para confiar na sua instância do ECR, e você deve configurar a autenticação para que o cluster use imagens Docker da AmazonECR. Para obter mais informações, consulte [Configuração YARN para acessar a Amazon ECR](#).

Para garantir que EMR os hosts da Amazon possam acessar as imagens armazenadas na AmazonECR, seu cluster deve ter as permissões da `AmazonEC2ContainerRegistryReadOnly` política associada ao perfil da instância. Para obter mais informações, consulte [Política AmazonEC2ContainerRegistryReadOnly](#).

Neste exemplo, o cluster deve ser criado com a seguinte configuração adicional para garantir que o ECR registro da Amazon seja confiável. Substitua o `123456789123.dkr.ecr.us-east-1.amazonaws.com` endpoint com seu ECR endpoint da 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": {}
}
]

```

## Usando PySpark com a Amazon ECR

O exemplo a seguir usa o PySpark Dockerfile, que será marcado e enviado para a Amazon. ECR. Depois de fazer o upload do Dockerfile, você pode executar o PySpark trabalho e consultar a imagem do Docker da Amazon. ECR

Depois de iniciar o cluster, use SSH para se conectar a um nó principal e executar os comandos a seguir para criar a imagem local do Docker a partir do exemplo do PySpark Dockerfile.

Primeiro, crie um diretório e um arquivo do Docker.

```

mkdir pyspark
vi pyspark/Dockerfile

```

Cole o conteúdo do PySpark Dockerfile e execute os comandos a seguir para criar uma imagem do Docker.

```

sudo docker build -t local/pyspark-example pyspark/

```

Crie o `emr-docker-examples` ECR repositório para os exemplos.

```

aws ecr create-repository --repository-name emr-docker-examples

```

Marque e faça o upload da imagem criada localmente para ECR, substituindo

*123456789123.dkr.ecr.us-east-1.amazonaws.com* com seu ECR endpoint.

```

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

```

Use SSH para se conectar ao nó primário e preparar um script Python com o nome do arquivo. `main.py` Cole o conteúdo a seguir no arquivo `main.py` e salve-o.

```
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)
```

Na Amazon EMR 6.0.0, para enviar o trabalho, faça referência ao nome da imagem do Docker. Defina os parâmetros de configuração adicionais para garantir que a execução do trabalho use o Docker como o runtime. Ao usar a Amazon ECR, é necessário referenciar o `config.json` arquivo contendo as credenciais usadas para se autenticar na 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
```

No Amazon EMR 6.1.0 e versões posteriores, para enviar o trabalho, faça referência ao nome da imagem do Docker. Quando a autenticação ECR automática estiver habilitada, execute o comando a seguir.

```
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
```

Quando o trabalho for concluído, anote o ID do YARN aplicativo e use o comando a seguir para obter a saída do PySpark trabalho.

```
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]]
```

## Usando o SparkR com a Amazon ECR

O exemplo a seguir usa o SparkR Dockerfile, que será marcado e enviado para ECR. Depois que o Dockerfile for carregado, você poderá executar o trabalho SparkR e consultar a imagem do Docker da Amazon ECR.

Depois de iniciar o cluster, use SSH para se conectar a um nó principal e executar os comandos a seguir para criar a imagem local do Docker a partir do exemplo do SparkR Dockerfile.

Primeiro, crie um diretório e o arquivo do Docker.

```
mkdir sparkr
vi sparkr/Dockerfile
```

Cole os conteúdos do arquivo do Docker do SparkR e execute os comandos a seguir para criar uma imagem do Docker.

```
sudo docker build -t local/sparkr-example sparkr/
```

Marque e faça o upload da imagem criada localmente para a Amazon ECR, substituindo *123456789123.dkr.ecr.us-east-1.amazonaws.com* com seu ECR endpoint da 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
```

Use SSH para se conectar ao nó primário e preparar um script R com o nome `sparkR.R`. Cole o conteúdo a seguir no arquivo `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()
```

Na Amazon EMR 6.0.0, para enviar o trabalho, consulte o nome da imagem do Docker. Defina os parâmetros de configuração adicionais para garantir que a execução do trabalho use o Docker como o runtime. Ao usar a Amazon ECR, eles `YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG` devem se referir ao `config.json` arquivo que contém as credenciais usadas para se autenticar. 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
```

No Amazon EMR 6.1.0 e versões posteriores, para enviar o trabalho, faça referência ao nome da imagem do Docker. Quando a autenticação ECR automática estiver habilitada, execute o seguinte comando.

```

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

```

Quando o trabalho for concluído, anote o ID do YARN aplicativo e use o comando a seguir para obter a saída do trabalho SparkR. Esse exemplo inclui testes para garantir que a randomForest biblioteca, a versão instalada e as notas de lançamento estejam disponíveis.

```

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:

```

## Use o AWS Glue Data Catalog como metastore para o Spark SQL

Usando a EMR versão 5.8.0 ou posterior da Amazon, você pode configurar o Spark SQL para usar o AWS Glue Data Catalog como metastore. Recomendamos essa configuração quando você precisa de um metastore persistente ou de um metastore compartilhado por diferentes clusters, serviços e aplicativos ou contas da AWS .

AWS O Glue é um serviço totalmente gerenciado de extração, transformação e carregamento (ETL) que torna simples e econômico categorizar seus dados, limpá-los, enriquecê-los e movê-los de forma confiável entre vários armazenamentos de dados. O AWS Glue Data Catalog fornece um repositório unificado de metadados em uma variedade de fontes e formatos de dados, com integração com a Amazon, Amazon EMR Redshift, RDS Redshift Spectrum, Athena e qualquer aplicativo compatível



com o metastore Apache Hive. AWS Os rastreadores do Glue podem inferir automaticamente o esquema dos dados de origem no Amazon S3 e armazenar os metadados associados no catálogo de dados. Para obter mais informações sobre o catálogo de dados, consulte [Preenchendo o catálogo de dados do AWS Glue](#) no Guia do desenvolvedor do AWS Glue.

Taxas separadas se aplicam ao AWS Glue. Há uma taxa mensal para armazenar e acessar os metadados no Catálogo de Dados, uma taxa horária cobrada por minuto pelas ETL tarefas do AWS Glue e pelo tempo de execução do rastreador, e uma taxa horária cobrada por minuto para cada endpoint de desenvolvimento provisionado. O Catálogo de Dados permite o armazenamento de até um milhão de objetos gratuitamente. Se você armazenar mais de um milhão de objetos, será cobrado USD \$1 por cada 100.000 objetos acima de um milhão. Um objeto no Catálogo de Dados é uma tabela, uma partição ou um banco de dados. Para obter mais informações, consulte [Definição de preço do Glue](#).

#### Important

Se você criou tabelas usando o Amazon Athena ou o Amazon Redshift Spectrum antes de 14 de agosto de 2017, bancos de dados e tabelas são armazenados em um catálogo gerenciado pelo Athena, que é separado do Glue Data Catalog. AWS Para integrar a Amazon EMR a essas tabelas, você deve fazer o upgrade para o AWS Glue Data Catalog. Para obter mais informações, consulte [Atualização para o catálogo de dados AWS Glue no Guia](#) do usuário do Amazon Athena.

## Especificando o AWS Glue Data Catalog como metastore

Você pode especificar o AWS Glue Data Catalog como metastore usando o AWS Management Console, AWS CLI, ou Amazon. EMR API Ao usar o CLI ou API, você usa a classificação de configuração do Spark para especificar o catálogo de dados. Além disso, com o Amazon EMR 5.16.0 e versões posteriores, você pode usar a classificação de configuração para especificar um catálogo de dados em outro. Conta da AWS Ao usar o console, você pode especificar o Catálogo de Dados usando Opções avançadas ou Opções rápidas.

#### Note

A opção de usar o AWS Glue Data Catalog também está disponível com o Zeppelin porque o Zeppelin é instalado com componentes do Spark. SQL

## Console

Para especificar o AWS Glue Data Catalog como a metastore do Spark com o novo console

1. Faça login no e abra AWS Management Console o EMR console da Amazon em <https://console.aws.amazon.com/emr>.
2. Em Amazon EMR on, EC2 no painel de navegação esquerdo, escolha Clusters e, em seguida, escolha Create cluster.
3. Em Pacote de aplicações, escolha Spark ou Personalizar. Se você personalizar o cluster, selecione Zeppelin ou Spark como uma das suas aplicações.
4. Em Configurações do Catálogo de Dados do AWS Glue, marque a caixa de seleção Usar para metadados de tabelas do Spark.
5. Escolha qualquer outra opção que se aplique ao cluster.
6. Para iniciar o cluster, escolha Criar cluster.

## AWS CLI

Para especificar o AWS Glue Data Catalog como o metastore do Spark com o AWS CLI

Para obter mais informações sobre como especificar uma classificação de configuração usando o AWS CLI e a Amazon EMRAPI, consulte [Configurar aplicações](#).

- Especifique o valor para `hive.metastore.client.factory.class` usando a classificação `spark-hive-site`, conforme mostrado neste exemplo:

```
[
 {
 "Classification": "spark-hive-site",
 "Properties": {
 "hive.metastore.client.factory.class":
 "com.amazonaws.glue.catalog.metastore.AWSGlueDataCatalogHiveClientFactory"
 }
 }
]
```

Para especificar um catálogo de dados em uma AWS conta diferente, adicione a `hive.metastore.glue.catalogid` propriedade conforme mostrado no exemplo a seguir. Substitua `acct-id` pela conta da AWS do Catálogo de Dados.

```
[
 {
 "Classification": "spark-hive-site",
 "Properties": {
 "hive.metastore.client.factory.class":
"com.amazonaws.glue.catalog.metastore.AWSGlueDataCatalogHiveClientFactory",
 "hive.metastore.glue.catalogid": "acct-id"
 }
 }
]
```

## IAMpermissões

O perfil da EC2 instância de um cluster deve ter IAM permissões para as ações do AWS Glue. Além disso, se você habilitar a criptografia para objetos do AWS Glue Data Catalog, a função também deverá ter permissão para criptografar, descriptografar e gerar o AWS KMS key usado para criptografia.

### Permissões para ações do AWS Glue

Se você usa o perfil de EC2 instância padrão da AmazonEMR, nenhuma ação é necessária. A política `AmazonElasticMapReduceforEC2Role` gerenciada anexada ao `EMR_EC2_DefaultRole` permite todas as ações necessárias do AWS Glue. No entanto, se você especificar um perfil de EC2 instância e permissões personalizados, deverá configurar as ações apropriadas do AWS Glue. Use a política gerenciada `AmazonElasticMapReduceforEC2Role` como ponto de partida. Para obter mais informações, consulte [Função de serviço para EC2 instâncias de cluster \(perfil de EC2 instância\)](#) no Amazon EMR Management Guide.

### Permissões para criptografar e descriptografar o Glue Data Catalog AWS

O perfil de instância precisa de permissão para criptografar e descriptografar dados usando a chave. Você não precisa configurar essas permissões se as duas instruções a seguir se aplicarem:

- Você ativa a criptografia para objetos do AWS Glue Data Catalog usando chaves gerenciadas para o AWS Glue.
- Você usa um cluster que está no mesmo catálogo Conta da AWS de dados do AWS Glue.

Caso contrário, você deverá adicionar a seguinte declaração à política de permissões anexada ao seu perfil de EC2 instância.

```
[
 {
 "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"
 }
]
 }
]
```

Para obter mais informações sobre a criptografia do AWS Glue Data Catalog, consulte [Criptografando seu catálogo de dados](#) no AWS Glue Developer Guide.

## Permissões baseadas em recursos

Se você usa o AWS Glue em conjunto com o Hive, o Spark ou o Presto na AmazonEMR, o AWS Glue oferece suporte a políticas baseadas em recursos para controlar o acesso aos recursos do Catálogo de Dados. Esses recursos incluem bancos de dados, tabelas, conexões e funções definidas pelo usuário. Para obter mais informações, consulte [Políticas baseadas em recursos no AWS Glue](#) no Guia do desenvolvedor do AWS Glue.

Ao usar políticas baseadas em recursos para limitar o acesso ao AWS Glue na AmazonEMR, o principal que você especifica na política de permissões deve ser a função ARN associada ao perfil de EC2 instância especificado quando um cluster é criado. Por exemplo, para uma política baseada em recursos anexada a um catálogo, você pode especificar a função da função de serviço padrão

ARN para instâncias de cluster EC2, `EMR_EC2_DefaultRole` como oPrincipal, usando o formato mostrado no exemplo a seguir:

```
arn:aws:iam::acct-id:role/EMR_EC2_DefaultRole
```

A ferramenta `acct-id` pode ser diferente do ID da conta AWS Glue. Isso permite o acesso a partir de EMR clusters em contas diferentes. Você pode especificar várias entidades principais, cada uma de uma conta diferente.

## Considerações ao usar o Catálogo de Dados do AWS Glue

Considere os seguintes itens ao usar o AWS Glue Data Catalog como metastore com o Spark:

- Ter um banco de dados padrão sem um local URI causa falhas quando você cria uma tabela. Como alternativa, use a cláusula `LOCATION` para especificar o local do bucket, como `s3://EXAMPLE-DOC-BUCKET`, ao usar `CREATE TABLE`. Ou crie tabelas em um banco de dados que não seja o banco de dados padrão.
- Não há suporte para renomear tabelas de dentro do AWS Glue.
- Quando você cria uma tabela do Hive sem especificar um `LOCATION`, os dados da tabela são armazenados no local especificado pela propriedade `hive.metastore.warehouse.dir`. Por padrão, esse é um local em HDFS. Se outro cluster precisar acessar a tabela, ele apresentará falha, a menos que tenha permissões adequadas para o cluster que criou a tabela. Além disso, como o HDFS armazenamento é transitório, se o cluster for encerrado, os dados da tabela serão perdidos e a tabela deverá ser recriada. Recomendamos que você especifique a `LOCATION` no Amazon S3 ao criar uma tabela do Hive usando o Glue. Como alternativa, você pode usar a classificação de configuração `hive-site` para especificar um local no Amazon S3 para `hive.metastore.warehouse.dir`, que se aplica a todas as tabelas do Hive. Se uma tabela for criada em um HDFS local e o cluster que a criou ainda estiver em execução, você poderá atualizar a localização da tabela para o Amazon S3 a partir do AWS Glue. Para obter mais informações, consulte Como [trabalhar com tabelas no console do AWS Glue](#) no Guia do desenvolvedor do AWS Glue.
- Valores de partição contendo aspas e apóstrofes não são compatíveis, por exemplo, `PARTITION (owner="Doe 's")`.
- [As estatísticas de coluna](#) são têm suporte para o emr-5.31.0 e versões posteriores.
- O uso da [autorização do Hive](#) não é compatível. Como alternativa, considere usar as [Políticas baseadas em recursos do AWS Glue](#). Para obter mais informações, consulte [Usar políticas baseadas em recursos para o catálogo de dados Amazon EMR Access to AWS Glue](#).

# Configurar o Spark

Você pode configurar o [Spark na Amazon EMR](#) com classificações de configuração. Para obter mais informações sobre classificações de configuração, consulte [Configurar aplicações](#).

As classificações de configuração do Spark na Amazon EMR incluem o seguinte:

- **spark**: define a propriedade `maximizeResourceAllocation` como verdadeira ou falsa. Quando verdadeiro, a Amazon configura EMR automaticamente as `spark-defaults` propriedades com base na configuração do hardware do cluster. Para obter mais informações, consulte [Utilizar o maximizeResourceAllocation](#).
- **spark-defaults**: define valores no arquivo `spark-defaults.conf`. Para obter mais informações, consulte [Configuração do Spark](#) na documentação do Spark.
- **spark-env**: define valores no arquivo `spark-env.sh`. Para obter mais informações, consulte [Variáveis de ambiente](#) na documentação do Spark.
- **spark-hive-site**: define os valores no `hive-site.xml` para o Spark.
- **spark-log4j**— (Amazon EMR lança 6.7.x e versões anteriores) Define valores no `log4j.properties` arquivo. Para obter mais informações, consulte o arquivo [log4j.properties.template](#) no GitHub.
- **spark-log4j2**— (Amazon EMR lança 6.8.0 e versões posteriores) Define valores no `log4j2.properties` arquivo. Para obter mais informações, consulte o arquivo [log4j2.properties.template](#) no GitHub.
- **spark-metrics**: define valores no arquivo `metrics.properties`. Para obter configurações e mais informações, consulte o arquivo [metrics.properties.template](#) no Github e as [Métricas](#) na documentação do Spark.

## Note

Se você estiver migrando cargas de trabalho do Spark para a Amazon EMR de outra plataforma, recomendamos que você teste suas cargas de trabalho com o [Padrões do Spark definidos pela Amazon EMR](#) antes de adicionar configurações personalizadas. A maioria dos clientes observa uma melhor performance com nossas configurações padrão.

## Tópicos

- [Padrões do Spark definidos pela Amazon EMR](#)
- [Configurando a coleta de lixo do Spark na Amazon 6.1.0 EMR](#)
- [Utilizar o maximizeResourceAllocation](#)
- [Configurar o comportamento de desativação de nós](#)
- [Variável de ThriftServer ambiente Spark](#)
- [Alterar as configurações padrão do Spark](#)
- [Migrar do Apache Log4j 1.x para Log4j 2.x](#)

## Padrões do Spark definidos pela Amazon EMR

A tabela a seguir mostra como a Amazon EMR define valores padrão spark-defaults que afetam os aplicativos.


### Padrões do Spark definidos pela Amazon EMR

| Configuração                                 | Descrição                                                                                                                                                                       | Valor padrão                                                                          |
|----------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|
| <code>spark.executor.memory</code>           | A quantidade de memória a ser usada por processo de executor. Por exemplo: 1g, 2g.                                                                                              | Essa configuração é determinada pelos tipos de instância core e de tarefa no cluster. |
| <code>spark.executor.cores</code>            | O número de núcleos para uso em cada executor.                                                                                                                                  | Essa configuração é determinada pelos tipos de instância core e de tarefa no cluster. |
| <code>spark.dynamicAllocation.enabled</code> | Quando verdadeira, use alocação dinâmica de recursos para aumentar e diminuir a escala verticalmente do número de executores registrados em uma aplicação com base na workload. | <code>true</code> (com Amazon EMR 4.4.0 e superior)                                   |

**Note**

O serviço Spark shuffle é configurado automaticamente pela Amazon. EMR

| Configuração                                                             | Descrição                                                                                                          | Valor padrão      |
|--------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------|-------------------|
| <code>spark.sql.hive.advancedPartitionPredicatePushdown.enabled</code>   | Quando verdadeiro, o pushdown avançado de predicados de partição para o metastore do Hive é habilitado.            | <code>true</code> |
| <code>spark.sql.hive.stringLikePartitionPredicatePushdown.enabled</code> | Envia os filtros <code>startsWith</code> , <code>contains</code> e <code>endsWith</code> para o metastore do Hive. | <code>true</code> |

 **Note**

O Glue não é compatível com o pushdown de predicados para `startsWith`, `contains` ou `endsWith`. Se você estiver usando o metastore do Glue e encontrar erros devido ao pushdown de predicados para essas funções, defina essa configuração como `false`.

## Configurando a coleta de lixo do Spark na Amazon 6.1.0 EMR

Definir configurações personalizadas de coleta de lixo com `spark.driver.extraJavaOptions` e `spark.executor.extraJavaOptions` resulta em falha na inicialização do driver ou do executor com o Amazon EMR 6.1 devido a uma configuração de coleta de lixo conflitante com o Amazon 6.1.0. EMR Para o Amazon EMR 6.1.0, a configuração padrão de coleta de lixo é definida por



meio de `spark.driver.defaultJavaOptions` e `spark.executor.defaultJavaOptions`. Essa configuração se aplica somente à Amazon EMR 6.1.0. JVMopções não relacionadas à coleta de lixo, como aquelas para configurar logging (`-verbose: class`), ainda podem ser definidas. `extraJavaOptions` Para obter mais informações, consulte [Propriedades das aplicações do Spark](#).

## Utilizar o `maximizeResourceAllocation`

Para configurar os executores para usarem o máximo de recursos possível em cada nó em um cluster, defina `maximizeResourceAllocation` como `true` na classificação de configuração do spark. Isso `maximizeResourceAllocation` é específico para a AmazonEMR. Quando você ativa `maximizeResourceAllocation`, a Amazon EMR calcula o máximo de recursos computacionais e de memória disponíveis para um executor em uma instância no grupo de instâncias principal. Em seguida, ele define as configurações `spark-defaults` correspondentes com base nos valores máximos calculados.

A Amazon EMR calcula o máximo de recursos computacionais e de memória disponíveis para um executor com base em um tipo de instância da frota principal de instâncias. Como cada frota de instâncias pode ter diferentes tipos e tamanhos de instância dentro de uma frota, a configuração do executor que a Amazon EMR usa pode não ser a melhor para seus clusters, por isso não recomendamos usar as configurações padrão ao usar a alocação máxima de recursos. Defina configurações personalizadas para seus clusters de frota de instâncias.

### Note

Você não deve usar a `maximizeResourceAllocation` opção em clusters com outros aplicativos distribuídos, como HBase o. A Amazon EMR usa YARN configurações personalizadas para aplicativos distribuídos, que podem entrar em conflito `maximizeResourceAllocation` e fazer com que os aplicativos Spark falhem.

Veja a seguir um exemplo de classificação de configuração do Spark com `maximizeResourceAllocation` definido como `true`.

```
[
 {
 "Classification": "spark",
 "Properties": {
 "maximizeResourceAllocation": "true"
 }
 }
]
```

```
}
]
```

Configurações configuradas em **spark-defaults** quando **maximizeResourceAllocation** está habilitado

| Configuração              | Descrição                                                                                                                              | Valor                                                                                                                                                                                                                                                                                                                                                         |
|---------------------------|----------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| spark.default.parallelism | O número padrão de partições é RDDs retornado por transformações como join reduceByKey, e paralelize quando não definido pelo usuário. | O dobro do número de CPU núcleos disponíveis para YARN contêineres.                                                                                                                                                                                                                                                                                           |
| spark.driver.memory       | Quantidade de memória a ser usada para o processo do driver, ou seja, onde SparkContext é inicializado. (por exemplo, 1g, 2g).         | A configuração é definida com base nos tipos de instância do cluster. No entanto, como o aplicativo do driver Spark pode ser executado na instância primária ou em uma das instâncias principais (por exemplo, nos modos YARN cliente e cluster, respectivamente), isso é definido com base no menor dos tipos de instância nesses dois grupos de instâncias. |
| spark.executor.memory     | Quantidade de memória para uso por processo de executor. (por exemplo, 1g, 2g)                                                         | A configuração é definida com base nos tipos de instância core e de tarefa do cluster.                                                                                                                                                                                                                                                                        |
| spark.executor.cores      | O número de núcleos para uso em cada executor.                                                                                         | A configuração é definida com base nos tipos de instância core e de tarefa do cluster.                                                                                                                                                                                                                                                                        |
| spark.executor.instances  | O número de executores.                                                                                                                | A configuração é definida com base nos tipos de                                                                                                                                                                                                                                                                                                               |

| Configuração | Descrição | Valor                                                                                                                                                                      |
|--------------|-----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|              |           | instância core e de tarefa do cluster. Defina a menos que <code>spark.dynamicAllocation.enabled</code> seja explicitamente definido como <code>true</code> ao mesmo tempo. |

## Configurar o comportamento de desativação de nós

Com a EMR versão 5.9.0 e superior da Amazon, o Spark na Amazon EMR inclui um conjunto de recursos para ajudar a garantir que o Spark gerencie o encerramento de nós sem problemas devido a um redimensionamento manual ou a uma solicitação automática de política de escalabilidade. A Amazon EMR implementa um mecanismo de negação de listagem no Spark que é construído sobre o mecanismo de YARN desativação. Esse mecanismo ajuda a garantir que não haja novas tarefas programadas em um nó que esteja sendo desativado, permitindo ao mesmo tempo que tarefas que já estão em execução sejam concluídas. Além disso, há recursos para ajudar a recuperar trabalhos do Spark com mais rapidez se os shuffle blocks forem perdidos no encerramento de um nó. O processo de recálculo é acionado mais cedo e otimizado para recalcular com mais rapidez e menos tentativas de fase, e é possível evitar falhas nos trabalhos provenientes de falhas de busca que são causadas por shuffle blocks ausentes.

### Important

A `spark.decommissioning.timeout.threshold` configuração foi adicionada na EMR versão 5.11.0 da Amazon para melhorar a resiliência do Spark ao usar instâncias spot. Nas versões anteriores, quando um nó usa uma instância spot e a instância é encerrada devido ao preço da oferta, o Spark pode não ser capaz de lidar com o encerramento com tranquilidade. Os trabalhos podem falhar e os novos cálculos de shuffle podem levar um tempo significativo. Por esse motivo, recomendamos o uso da versão 5.11.0, ou posterior, se você usar instâncias spot.

## Configurações de desativação de nós do Spark

| Configuração                                         | Descrição                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | Valor padrão      |
|------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------|
| <code>spark.blacklist.decommissioning.enabled</code> | Quando definido como <code>true</code> , a negação do Spark lista os nós que estão no <code>decommissioning</code> estado em YARN. O Spark não programa novas tarefas em executores que estejam em execução nesse nó. É permitido que as tarefas já em execução sejam concluídas.                                                                                                                                                                                                                                                                                                                                                 | <code>true</code> |
| <code>spark.blacklist.decommissioning.timeout</code> | O período em que um nó no estado <code>decommissioning</code> permanece na lista de negação. Por padrão, esse valor é definido como uma hora, que também é o padrão para <code>yarn.resourcemanager.decommissioning.timeout</code> . Para garantir que um nó permaneça na lista de negação por todo o período de desativação, defina esse valor como igual ou maior do que <code>yarn.resourcemanager.decommissioning.timeout</code> . Depois que o tempo limite de desativação expirar, o nó passa para um estado <code>decommissioned</code> e a Amazon EMR pode encerrar a instância do nó. EC2 Se alguma tarefa ainda estiver | 1h                |

| Configuração | Descrição                                                                                                                     | Valor padrão |
|--------------|-------------------------------------------------------------------------------------------------------------------------------|--------------|
|              | em execução após o tempo limite expirar, ela será perdida ou eliminada e reprogramada em executores executados em outros nós. |              |

| Configuração                                         | Descrição                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | Valor padrão |
|------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------|
| <code>spark.decommissioning.timeout.threshold</code> | Disponível na EMR versão 5.11.0 ou posterior da Amazon. Especificado em segundos. Quando um nó passa para o estado de descomissionamento, se o host for desativado dentro de um período de tempo igual ou menor que esse valor, a Amazon EMR não apenas lista o nó, mas também limpa o estado do host (conforme especificado por <code>spark.resourceManager.cleanupExpiredHost</code> ) sem esperar que o nó faça a transição para um estado descomissionado. Isso permite que o Spark lide melhor com o encerramento de instâncias spot, pois as instâncias spot são desativadas dentro de um tempo de espera de 20 segundos, independentemente do valor do <code>yarn.resourcemanager.decommissioning.timeout</code> , o que pode não dar aos outros nós tempo suficiente e para que eles leiam arquivos embaralhados. | 20s          |

| Configuração                                                      | Descrição                                                                                                                                                                                                                                                                                                                        | Valor padrão      |
|-------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------|
| <code>spark.resourceManager.cleanupExpiredHost</code>             | Quando definido como <code>true</code> , o Spark cancela o registro de todos os dados em cache e os shuffle blocks armazenados nos executores nos nós que estejam no estado <code>decommissioned</code> . Isso acelera o processo de recuperação.                                                                                | <code>true</code> |
| <code>spark.stage.attempt.ignoreOnDecommissionFetchFailure</code> | Quando definido como <code>true</code> , ajuda a evitar falhas do Spark nas fases e a falha no trabalho devido ao excesso de falhas de busca dos nós desativados. As buscas com falha de shuffle blocks de nó no estado <code>decommissioned</code> não são contabilizadas para o número máximo de falhas de busca consecutivas. | <code>true</code> |

## Variável de ThriftServer ambiente Spark

O Spark define a variável de ambiente da Porta do servidor Thrift do Hive, `HIVE_SERVER2_THRIFT_PORT`, como 10001.

## Alterar as configurações padrão do Spark

Você altera os padrões em `spark-defaults.conf` usando a classificação de configuração `spark-defaults` ou a configuração `maximizeResourceAllocation` na classificação de configuração `spark`.

Os procedimentos a seguir mostram como modificar as configurações usando o console CLI ou.

## Para criar um cluster com spark.executor.memory definido como 2g usando o CLI

- Crie um cluster com o Spark instalado e com spark.executor.memory definido como 2g usando o comando a seguir, que faz referência a um arquivo, myConfig.json, armazenado no 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

Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

### myConfig.json:

```
[
 {
 "Classification": "spark-defaults",
 "Properties": {
 "spark.executor.memory": "2G"
 }
 }
]
```

## Criar um cluster com spark.executor.memory definido como 2g usando o console

1. Navegue até o novo EMR console da Amazon e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte [Usar o console antigo](#).
2. Escolha Create cluster (Criar cluster), Go to advanced options (Ir para opções avançadas).
3. Escolha Spark.
4. Em Edit software settings (Editar configurações de software), deixe Enter configuration (Inserir configuração) selecionado e insira a seguinte configuração:



```
classification=spark-defaults,properties=[spark.executor.memory=2G]
```

5. Selecione outras opções, escolha `eCreate cluster` (Criar cluster).

Para definir `maximizeResourceAllocation`

- Crie um cluster com o Spark instalado e `maximizeResourceAllocation` definido como verdadeiro usando o AWS CLI, referenciando um arquivo `myConfig.json`, armazenado no 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

Os caracteres de continuação de linha do Linux (\) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (^).

`myConfig.json`:

```
[
 {
 "Classification": "spark",
 "Properties": {
 "maximizeResourceAllocation": "true"
 }
 }
]
```

#### Note

Com a Amazon EMR versão 5.21.0 e posterior, você pode substituir as configurações do cluster e especificar classificações de configuração adicionais para cada grupo de instâncias

em um cluster em execução. Você faz isso usando o EMR console da Amazon, o AWS Command Line Interface (AWS CLI) ou AWS SDK o. Para obter mais informações, consulte [Supplying a Configuration for an Instance Group in a Running Cluster](#).

## Migrar do Apache Log4j 1.x para Log4j 2.x

As versões 3.2.x e anteriores do [Apache Spark](#) usam o Apache Log4j 1.x herdado e o arquivo `log4j.properties` para configurar o Log4j nos processos do Spark. As versões 3.3.0 e posteriores do Apache Spark usam o Apache Log4j 2.x e o arquivo `log4j2.properties` para configurar o Log4j nos processos do Spark.

Se você configurou o Apache Spark Log4j usando uma EMR versão da Amazon inferior à 6.8.0, então você deve remover a classificação de `spark-log4j` configuração legada e migrar para a classificação de configuração e formato de chave antes de poder atualizar para a `spark-log4j2` Amazon 6.8.0 ou posterior. EMR A `spark-log4j` classificação legada faz com que a criação do cluster falhe com um `ValidationException` erro nas EMR versões 6.8.0 e posteriores da Amazon. Você não será cobrado por uma falha relacionada à incompatibilidade do Log4j, mas deverá remover a classificação de configuração `spark-log4j` extinta para continuar.

Para obter mais informações sobre a migração do Apache Log4j 1.x para o Log4j 2.x, consulte o [Guia de migração do Apache Log4j](#) e o [Modelo do Spark Log4j 2](#) no GitHub.

### Note

Com a AmazonEMR, o Apache Spark usa um `log4j2.properties` arquivo em vez do `arquivo.xml` descrito no Guia de migração do [Apache Log4j](#). Além disso, não recomendamos o uso do método de ponte do Log4j 1.x para conversão para Log4j 2.x.

## Otimizar a performance do Spark

A Amazon EMR fornece vários recursos de otimização de desempenho para o Spark. Este tópico explica detalhadamente cada recurso de otimização.

Para obter mais informações sobre como definir a configuração do Spark, consulte [Configurar o Spark](#).

## Execução adaptável de consultas

A execução adaptável de consultas é uma estrutura para a reotimização de planos de consulta com base nas estatísticas de runtime. A partir do Amazon EMR 5.30.0, as seguintes otimizações adaptativas de execução de consultas do Apache Spark 3 estão disponíveis no Apache Amazon Runtime for Spark 2. EMR

- Conversão adaptável de junções
- Aglutinação adaptável de partições aleatórias

### Conversão adaptável de junções

A conversão adaptativa de junções melhora o desempenho da consulta ao converter sort-merge-join operações em broadcast-hash-joins operações com base nos tamanhos de tempo de execução dos estágios da consulta. Broadcast-hash-joins tendem a ter um desempenho melhor quando um lado da junção é pequeno o suficiente para transmitir com eficiência sua saída para todos os executores, evitando assim a necessidade de trocar aleatoriamente e classificar os dois lados da junção. A conversão adaptativa de junções amplia a variedade de casos em que o Spark funciona automaticamente. broadcast-hash-joins

Esse recurso está habilitado por padrão. Ele pode ser desabilitado pela definição de `spark.sql.adaptive.enabled` como `false`, o que também desabilita a estrutura adaptável de execução de consultas. O Spark decide converter a em a broadcast-hash-join quando sort-merge-join a estatística do tamanho do tempo de execução de um dos lados da junção não excede `spark.sql.autoBroadcastJoinThreshold`, cujo padrão é 10.485.760 bytes (10 MiB).

### Aglutinação adaptável de partições aleatórias

A aglutinação adaptável de partições aleatórias melhora a performance das consultas ao aglutinar pequenas partições aleatórias contíguas para evitar a sobrecarga de ter muitas tarefas pequenas. Isso permite que você configure antecipadamente um número maior de partições aleatórias iniciais que, em seguida, são reduzidas em runtime para um tamanho desejado, aumentando as probabilidades da existência de partições aleatórias distribuídas mais uniformemente.

Esse atributo é habilitado por padrão, a menos que `spark.sql.shuffle.partitions` seja definido explicitamente. Ele pode ser habilitado pela definição de `spark.sql.adaptive.coalescePartitions.enabled` como `true`. O número inicial de partições aleatórias e o tamanho da partição de destino podem ser ajustados com o uso das propriedades `spark.sql.adaptive.coalescePartitions.minPartitionNum` e

`spark.sql.adaptive.advisoryPartitionSizeInBytes`, respectivamente. Consulte a tabela a seguir para obter mais detalhes sobre as propriedades do Spark relacionadas a esse atributo.

### Propriedades da partição de aglutinação adaptável do Spark

| Propriedade                                                        | Valor padrão                                                                                   | Descrição                                                                                                                                                                                                                                                                           |
|--------------------------------------------------------------------|------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>spark.sql.adaptive.coalescePartitions.enabled</code>         | verdadeiro, a menos que <code>spark.sql.shuffle.partitions</code> seja explicitamente definido | Quando verdadeiro e <code>spark.sql.adaptive.enabled</code> é verdadeiro, o Spark aglutina partições aleatórias contíguas de acordo com o tamanho de destino (especificado por <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code> ), para evitar muitas tarefas pequenas. |
| <code>spark.sql.adaptive.advisoryPartitionSizeInBytes</code>       | 64 MB                                                                                          | O tamanho do comunicado, em bytes, da partição aleatória durante a aglutinação. Essa configuração só tem efeito quando <code>spark.sql.adaptive.enabled</code> e <code>spark.sql.adaptive.coalescePartitions.enabled</code> são ambos <code>true</code> .                           |
| <code>spark.sql.adaptive.coalescePartitions.minPartitionNum</code> | 25                                                                                             | O número mínimo de partições aleatórias após a aglutinação. Essa configuração só tem efeito quando <code>spark.sql.adaptive.enabled</code> e <code>spark.sql.adaptive.coalescePartitions.enabled</code> são ambos <code>true</code> .                                               |

| Propriedade                                                            | Valor padrão | Descrição                                                                                                                                                                                                                                |
|------------------------------------------------------------------------|--------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>spark.sql.adaptive.coalescePartitions.initialPartitionNum</code> | 1000         | O número inicial de partições aleatórias antes da aglutinação. Essa configuração só tem efeito quando <code>spark.sql.adaptive.enabled</code> e <code>spark.sql.adaptive.coalescePartitions.enabled</code> são ambos <code>true</code> . |

## Corte de partição dinâmico

O corte de partição dinâmico melhora o desempenho da tarefa selecionando as partições específicas com mais precisão em uma tabela que precisa ser lida e processada em uma consulta específica. Reduzindo a quantidade de dados que são lidos e processados, um tempo significativo é economizado na execução da tarefa. Com o Amazon EMR 5.26.0, esse recurso é ativado por padrão. Com o Amazon EMR 5.24.0 e 5.25.0, você pode ativar esse recurso definindo a propriedade Spark de dentro `spark.sql.dynamicPartitionPruning.enabled` do Spark ou ao criar clusters.

### Propriedades do corte de partição dinâmico do Spark

| Propriedade                                                                    | Valor padrão      | Descrição                                                                                                                                                                                                       |
|--------------------------------------------------------------------------------|-------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>spark.sql.dynamicPartitionPruning.enabled</code>                         | <code>true</code> | Quando verdadeiro, habilite o corte de partição dinâmico.                                                                                                                                                       |
| <code>spark.sql.optimizer.dynamicPartitionPruning.enforceBroadcastReuse</code> | <code>true</code> | Quando <code>true</code> , o Spark conduz uma verificação defensiva antes da execução da consulta para garantir que a reutilização de trocas de transmissão em filtros de corte dinâmico não seja interrompida. |

| Propriedade | Valor padrão | Descrição                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|-------------|--------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|             |              | <p>ida por regras de preparação o posteriores, como regras colunares definidas pelo usuário. Quando a reutilização é interrompida e essa configuração é <code>true</code>, o Spark remove os filtros de corte dinâmico afetados para evitar problemas de performance e correção. Podem surgir problemas de correção quando a troca de transmissão do filtro de corte dinâmico produz resultados diferentes e inconsistentes da troca de transmissão da operação de junção correspondente. A definição dessa configuração como <code>false</code> deve ser feita com cuidado; ela permite contornar cenários, como quando a reutilização é interrompida por regras colunares definidas pelo usuário. Quando a execução adaptável de consultas está habilitada, a reutilização da transmissão sempre é imposta.</p> |

Essa otimização melhora as capacidades existentes do Spark 2.4.2, que tem suporte apenas à redução de predicados estáticos que podem ser resolvidos na hora do planejamento.

Veja a seguir exemplos de redução de predicados estáticos no Spark 2.4.2.

```
partition_col = 5

partition_col IN (1,3,5)

partition_col between 1 and 3

partition_col = 1 + 3
```

O corte de partição dinâmico permite que o mecanismo do Spark deduza dinamicamente no runtime quais partições precisam ser lidas e quais podem ser eliminadas com segurança. Por exemplo, a consulta a seguir envolve duas tabelas: a tabela `store_sales` contém as vendas totais de todas as lojas e é particionada por região, e a tabela `store_regions` contém um mapeamento de regiões por país. As tabelas contêm dados sobre as lojas que estão distribuídas pelo mundo, mas estão consultado dados somente para a América do Norte.

```
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'
```

Sem o corte de partição dinâmico, essa consulta lerá todas as regiões antes de filtrar o subconjunto de regiões que corresponde aos resultados da subconsulta. Com o corte de partição dinâmico, essa consulta lerá e processará apenas as partições para as regiões retornadas na subconsulta. Isso economiza tempo e recursos ao ler menos dados do armazenamento e processar menos registros.

## Nivelamento de subconsultas escalares

Esta otimização melhora o desempenho de consultas que têm subconsultas escalares na mesma tabela. Com o Amazon EMR 5.26.0, esse recurso é ativado por padrão. Com o Amazon EMR 5.24.0 e 5.25.0, você pode habilitá-lo definindo a propriedade Spark de dentro `spark.sql.optimizer.flattenScalarSubqueriesWithAggregates.enabled` do Spark ou ao criar clusters. Quando essa propriedade é definida como `true`, o otimizador de consultas nivela as subconsultas escalares que usam a mesma relação, quando possível. As subconsultas escalares são niveladas enviando qualquer predicado presente na subconsulta para as funções agregadas e executando uma agregação, com todas as funções agregadas, por relação.

Veja a seguir um exemplo de consulta que se beneficia dessa otimização.

```
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

```

A otimização reescreve a consulta anterior como:

```

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);

```

Observe que a consulta regravada lê a tabela do estudante somente uma vez, e os predicados das três subconsultas são enviados para a função avg.

## DISTINCTantes INTERSECT

Essa otimização otimiza as junções durante o uso. INTERSECT Com o Amazon EMR 5.26.0, esse recurso é ativado por padrão. Com o Amazon EMR 5.24.0 e 5.25.0, você pode habilitá-lo definindo a propriedade Spark de dentro `spark.sql.optimizer.distinctBeforeIntersect.enabled` do Spark ou ao criar clusters. As consultas usadas INTERSECT são automaticamente convertidas para usar uma junção semi esquerda. Quando essa propriedade é definida como verdadeira, o otimizador de consultas envia o DISTINCT operador para os filhos de INTERSECT se detectar que o DISTINCT operador pode fazer com que a junção semi esquerda seja a em vez de a. BroadcastHashJoin SortMergeJoin

Veja a seguir um exemplo de consulta que se beneficia dessa otimização.

```

(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)

```

Sem habilitar essa propriedade `spark.sql.optimizer.distinctBeforeIntersect.enabled`, a consulta será regravada do modo a seguir.

```

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
```

Quando você habilita essa propriedade

`spark.sql.optimizer.distinctBeforeIntersect.enabled`, a consulta pode ser regravada do modo a seguir.

```
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
```

## Junção com filtro Bloom

Essa otimização pode melhorar o desempenho de algumas junções ao fazer uma pré-filtragem de um lado de uma junção usando um [filtro Bloom](#) gerado pelos valores do outro lado da junção. Com o Amazon EMR 5.26.0, esse recurso é ativado por padrão. Com o Amazon EMR 5.25.0, você pode ativar esse recurso definindo a propriedade Spark como `true` de dentro do Spark ou `spark.sql.bloomFilterJoin.enabled` ao criar clusters.

Veja a seguir um exemplo de consulta que podem se beneficiar de um filtro Bloom.

```
select count(*)
from sales, item
where sales.item_id = item.id
and item.category in (1, 10, 16)
```

Quando esse recurso está habilitado, o filtro Bloom é criado de todos os IDs de itens cuja categoria está no conjunto de categorias que está sendo consultado. Ao verificar a tabela de vendas, o filtro Bloom é usado para determinar quais são as vendas de itens que definitivamente não estão no conjunto definido pelo filtro Bloom. Assim, o filtro pode excluir essas vendas identificadas o mais cedo possível.

## Reordenação de junção otimizada

Essa otimização pode melhorar o desempenho de consultas reordenando junções que envolvem tabelas com filtros. Com o Amazon EMR 5.26.0, esse recurso é ativado por padrão. Com o Amazon EMR 5.25.0, você pode ativar esse recurso definindo o parâmetro `spark.sql.optimizer.sizeBasedJoinReorder.enabled` de configuração do Spark como verdadeiro. O comportamento padrão no Spark é unir tabelas da esquerda para a direita, conforme listado na consulta. Essa estratégia pode perder oportunidades de executar junções menores com filtros primeiro a fim de beneficiar junções mais caras posteriormente.

O exemplo de consulta a seguir relata todos os itens devolvidos de todas as lojas de um país. Sem a reordenação de junção otimizada, o Spark une as duas tabelas grandes `store_sales` e `store_returns` primeiro, depois, as une a `store` e, por fim, a `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'
```

Com reordenação de junção otimizada o Spark une `store_sales` a `store` primeiro, pois `store` tem um filtro e é menor que `store_returns` e `broadcastable`. Em seguida, o Spark une a `store_returns` e, por fim, a `item`. Se `item` tivesse um filtro e fosse possível de ser transmitido, ele também se qualificaria para a reordenação, resultando em `store_sales` se unindo a `store`, depois a `item` e, por fim, a `store_returns`.

## Armazenamento em Cache de Fragmentos Resultantes do Spark

O Amazon EMR 6.6.0 e versões posteriores incluem o recurso opcional Spark Result Fragment Caching, que armazena automaticamente fragmentos de resultados em cache. Esses fragmentos resultantes são partes dos resultados de subárvores de consultas que são armazenadas em um bucket do Amazon S3 de sua escolha. Os fragmentos resultantes da consulta armazenados são reutilizados em execuções de consultas subsequentes, resultando em consultas mais rápidas.

O Result Fragment Caching analisa suas SQL consultas do Spark e armazena em cache os fragmentos de resultados elegíveis no local especificado do S3. Em execuções subsequentes da consulta, os fragmentos resultantes utilizáveis da consulta são automaticamente detectados e obtidos no S3. O Armazenamento em Cache de Fragmentos Resultantes difere do Armazenamento em Cache de Conjuntos Resultantes, em que as consultas subsequentes precisam corresponder

exatamente à consulta original para retornar os resultados do cache. Quando usado para consultas repetidamente direcionadas a um subconjunto estático dos dados, o Armazenamento em Cache de Fragmentos Resultantes acelera significativamente a performance.

Considere a consulta a seguir, que conta pedidos até o ano de 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
```

Com o passar do tempo, essa consulta precisa ser executada todos os dias para informar o total de vendas do ano. Sem o Armazenamento em Cache de Fragmentos Resultantes, os resultados de todos os dias do ano precisam ser recalculados todos os dias. A consulta ficará mais lenta com o tempo e será mais lenta no final do ano, quando todos os 365 dias de resultados precisarão ser recalculados.

Ao ativar o Armazenamento em Cache de Fragmentos Resultantes, você usa os resultados de todos os dias anteriores do ano com base no cache. A cada dia, o atributo deve recalculá apenas um dia de resultados. Depois que o atributo computa o fragmento resultante, o atributo armazena o fragmento em cache. Como resultado, os tempos de consultas habilitadas para cache são curtos e permanecem constantes para cada consulta subsequente.

## Habilitar o Armazenamento em Cache de Fragmentos Resultantes do Spark

Para habilitar o Armazenamento em Cache de Fragmentos Resultantes do Spark, execute as seguintes etapas:

1. Crie um bucket de cache no Amazon S3 e autorize o acesso de leitura/gravação para EMRFS. Para obter mais informações, consulte [Autorizando o acesso aos EMRFS dados no Amazon S3](#).
2. Defina a configuração EMR do Amazon Spark para ativar o recurso.

```
spark.subResultCache.enabled = true
spark.subResultCache.fs.root.path = s3://DOC-EXAMPLE-BUCKET/cache_dir/
```

3. Habilite o gerenciamento do ciclo de vida do S3 para que o bucket limpe automaticamente os arquivos de cache.
4. Opcionalmente, configure as `maxBufferSize` propriedades `reductionRatioThreshold` e para ajustar ainda mais o recurso.

```
spark.sql.subResultCache.reductionRatioThreshold
spark.sql.subResultCache.maxBufferSize
```

## Considerações ao usar o Armazenamento em Cache de Fragmentos Resultantes

A economia de custos obtida quando você usa resultados já armazenados em cache no Amazon S3 em vez de recalculá-los aumenta com o número de vezes que os mesmos resultados armazenados em cache podem ser usados. Consultas com verificações de grandes tabelas seguidas por filtros ou agregações de hash que reduzem o tamanho do resultado em um fator de pelo menos 8 (ou seja, uma proporção de pelo menos 8:1 em tamanho de entrada:resultados) serão as que mais se beneficiarão desse atributo. Quanto maior a taxa de redução entre a entrada e os resultados, maior a relação custo-benefício. Consultas com taxas de redução menores, mas que contêm etapas computacionais caras entre a verificação da tabela e o filtro ou as agregações, também serão beneficiadas, desde que o custo de produzir os resultados seja maior do que o custo de buscá-los no Amazon S3. Por padrão, o Armazenamento em Cache de Fragmentos Resultantes só entra em vigor quando detecta que uma taxa de redução será de pelo menos 8:1.

Quando suas consultas reutilizam repetidamente os resultados armazenados em cache, os benefícios desse atributo são maiores. Consultas em janelas contínuas e incrementais são bons exemplos. Por exemplo, uma consulta de janela contínua de 30 dias que já foi executada por 29 dias precisaria extrair apenas 1/30 dos dados de destino de sua fonte de entrada original e usaria fragmentos de resultados armazenados em cache nos 29 dias anteriores. Uma consulta de janela incremental se beneficiaria ainda mais, já que o início da janela permanece fixo: em cada invocação da consulta, uma porcentagem menor do processamento exigirá a leitura da fonte de entrada.

Veja a seguir considerações adicionais sobre o uso do Armazenamento em Cache de Fragmentos Resultantes:

- As consultas que não têm como alvo os mesmos dados com os mesmos fragmentos de consulta terão uma baixa taxa de acertos de cache e, portanto, não se beneficiarão desse atributo.
- Consultas com taxas de redução baixas que não contêm etapas de computação caras resultarão em resultados armazenados em cache que são aproximadamente tão caros de ler quanto de processar inicialmente.
- A primeira consulta sempre demonstrará uma pequena regressão devido ao custo de gravação no cache.
- O atributo Armazenamento em Cache de Fragmentos Resultantes funciona exclusivamente com arquivos Parquet. Arquivos em outros formatos não são compatíveis.
- Os buffers do atributo Armazenamento em Cache de Fragmentos Resultantes só tentarão armazenar em cache verificações com tamanhos de divisão de arquivos de 128 MB ou mais. Com a configuração padrão do Spark, o Armazenamento em Cache de Fragmentos Resultantes será desabilitado se o tamanho da verificação (tamanho total de todos os arquivos que estão sendo verificados) dividido pelo número de núcleos do executor for menor que 128 MB. Quando qualquer uma das configurações do Spark listadas abaixo for definida, o tamanho da divisão de arquivos será:

```
min(maxPartitionBytes, max(openCostInBytes, scan size / minPartitionNum))
```

- spark.sql. leafNodeDefaultParalelismo (o valor padrão é spark.default.parallelism)
- spark.sql.files. minPartitionNum (o valor padrão é spark.sql. leafNodeDefaultParalelismo)
- spark.sql.files. openCostInBytes
- spark.sql.files. maxPartitionBytes
- O recurso Result Fragment Caching armazena em cache a granularidade da RDD partição. A taxa de redução descrita anteriormente, cujo padrão é 8:1, é avaliada por partição. RDD Cargas de trabalho com taxas por RDD redução maiores e menores que 8:1 podem ter menores benefícios de desempenho do que cargas de trabalho com taxas por RDD redução que são consistentemente menores que 8:1.
- O recurso Result Fragment Caching usa um buffer de gravação de 16 MB por padrão para cada RDD partição que está sendo armazenada em cache. Se mais de 16 MB forem armazenados em cache por RDD partição, o custo de determinar que uma gravação não é possível pode resultar em uma regressão de desempenho.
- Embora, por padrão, o Result Fragment Caching não tente armazenar em cache os resultados da RDD partição com uma taxa de redução menor que 8:1 e limite seu buffer de gravação em 16 MB, esses dois valores podem ser ajustados por meio das seguintes configurações:

```
spark.sql.subResultCache.reductionRatioThreshold (default: 8.0)
spark.sql.subResultCache.maxBufferSize (default: 16MB, max: 64MB)
```

- Vários clusters usando a mesma EMR versão da Amazon podem compartilhar o mesmo local de cache. Para garantir a exatidão dos resultados, o Result Fragment Caching não usará resultados de cache escritos por diferentes versões da Amazon. EMR
- O cache de fragmentos de resultados será desativado automaticamente para casos de uso do Spark Streaming ou quando RecordServer o Apache Ranger for usado. AWS Lake Formation
- O resultado fragmenta o cache de leitura/gravação, o uso e os buckets do EMRFS Amazon S3. CSE/A SSE KMS criptografia SSE S3/ é suportada.

## Use o Nvidia RAPIDS Accelerator para Apache Spark

Com a EMR versão 6.2.0 e posterior da Amazon, você pode usar o plug-in [RAPIDS Accelerator for Apache Spark da Nvidia para](#) acelerar o Spark usando tipos de instância de unidade de processamento EC2 gráfico (). GPU RAPIDS Accelerator GPU acelerará seus pipelines de ciência de dados do Apache Spark 3.0 sem alterações no código e acelerará o processamento de dados e o treinamento de modelos, ao mesmo tempo em que reduzirá substancialmente os custos de infraestrutura.

As seções a seguir orientam você na configuração do EMR cluster para usar o RAPIDS plug-in Spark para Spark.

### Selecionar tipos de instância

Para usar o RAPIDS plug-in Nvidia Spark- para Spark, os grupos de instâncias principais e de tarefas devem usar tipos de EC2 GPU instância que atendam aos [requisitos de hardware](#) do Spark-. RAPIDS Para ver uma lista completa dos tipos de GPU instâncias EMR suportadas pela Amazon, consulte [Tipos de instância compatíveis](#) no Guia EMR de gerenciamento da Amazon. O tipo de instância para o grupo de instâncias primário pode ser um GPU ou não, mas os tipos de ARM instância não são compatíveis. GPU

### Definir as configurações da aplicação para o cluster

1. Permita EMR que a Amazon instale os plug-ins em seu novo cluster

Para instalar plug-ins, forneça a seguinte configuração ao criar o cluster:

```
{
 "Classification":"spark",
 "Properties":{
 "enableSparkRapids":"true"
 }
}
```

## 2. Configurar YARN para usar GPU

Para obter detalhes sobre como usar GPU onYARN, consulte [Usando GPU on YARN na documentação do Apache Hadoop](#). Os exemplos a seguir mostram exemplos de YARN configurações para as versões EMR 6.x e 7.x da Amazon:

### Amazon EMR 7.x

#### Exemplo YARN de configuração para 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

### Exemplo YARN de configuração para 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. Configurar o Spark para usar RAPIDS

Aqui estão as configurações necessárias para permitir que o Spark use RAPIDS o plug-in:

```

{
 "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- A biblioteca Spark](#) na XGBoost documentação também está disponível quando o RAPIDS plug-in Spark está habilitado em seu cluster. Você pode usar a seguinte configuração para se integrar XGBoost ao seu trabalho do Spark:

```

{
 "Classification":"spark-defaults",
 "Properties":{
 "spark.submit.pyFiles":"/usr/lib/spark/jars/xgboost4j-spark_3.0-1.4.2-0.3.0.jar"
 }
}

```

Para configurações adicionais do Spark que você pode usar para ajustar um EMR cluster GPU acelerado, consulte o guia de ajuste do [Rapids Accelerator para Apache Spark](#) na documentação do NVIDIA.GitHub.io.

### 4. Configurar o YARN Capacity Scheduler

DominantResourceCalculator deve ser configurado para permitir o GPU agendamento e o isolamento. Para obter mais informações, consulte [Usando GPU on YARN na documentação do Apache Hadoop](#).

```
{
 "Classification": "capacity-scheduler",
 "Properties": {
 "yarn.scheduler.capacity.resource-calculator": "org.apache.hadoop.yarn.util.resource.DominantResourceCalculator"
 }
}
```

## 5. Crie um JSON arquivo para incluir suas configurações

Você pode criar um JSON arquivo que contenha sua configuração para usar o RAPIDS plug-in em seu cluster Spark. Você fornece o arquivo posteriormente ao executar o cluster.

Você pode armazenar o arquivo localmente ou no S3. Para obter mais informações sobre como fornecer configurações de aplicações para os clusters, consulte [Configurar aplicações](#).

Use os exemplos de arquivos a seguir como modelos para criar suas próprias configurações.

### Amazon EMR 7.x

#### **my-configurations.json** Arquivo de exemplo para 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** Arquivo de exemplo para 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"
 }
 }
]
]
```

## Adicionar uma ação de bootstrap para o cluster

Para obter mais informações sobre como fornecer scripts de ação de bootstrap ao criar seu cluster, consulte [Noções básicas sobre ações de bootstrap](#) no Amazon EMR Management Guide.

Os scripts de exemplo a seguir mostram como criar um arquivo de ação de bootstrap para Amazon EMR 6.x e 7.x:

### Amazon EMR 7.x

#### **my-bootstrap-action.sh** Arquivo de exemplo para Amazon EMR 7.x

Para usar YARN para gerenciar GPU recursos com as versões do Amazon EMR 7.x, você deve montar manualmente CGroup v1 em seu cluster. Isso pode ser feito com um script de ação de bootstrap, conforme mostrado neste exemplo.

```
#!/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+rxw -R /spark-rapids-cgroup
```

### Amazon EMR 6.x

#### **my-bootstrap-action.sh** Arquivo de exemplo para Amazon EMR 6.x

Para versões do Amazon EMR 6.x, você deve abrir CGroup permissões YARN em seu cluster. Isso pode ser feito com um script de ação de bootstrap, conforme mostrado neste exemplo.

```
#!/bin/bash
```

```
set -ex

sudo chmod a+rxw -R /sys/fs/cgroup/cpu,cpuacct
sudo chmod a+rxw -R /sys/fs/cgroup/devices
```

## Iniciar o cluster

A última etapa é iniciar o cluster com as configurações de cluster mencionadas acima. Aqui está um exemplo de comando para iniciar um cluster da Amazon EMRCLI:

```
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
```

## Acessar o shell do Spark

O shell do Spark é baseado no Scala REPL (Read-Eval-Print-Loop). Ele permite criar programas Spark interativamente e enviar trabalhos à estrutura. Você pode acessar o shell do Spark conectando-se ao nó primário com SSH e invocando `spark-shell`. Para obter mais informações sobre a conexão com o nó primário, consulte [Conecte-se ao nó primário usando SSH](#) o Amazon EMR Management Guide. Os exemplos a seguir usam os logs de acesso HTTP do Apache Server armazenados no Amazon S3.

### Note

O bucket nesses exemplos está disponível para clientes que podem acessar a região Leste dos EUA (Norte da Virgínia).

Por padrão, o shell do Spark cria seu próprio [SparkContext](#) objeto chamado `sc`. Você pode usar esse contexto se for necessário dentro do REPL. `sqlContext` também está disponível na concha e é uma [HiveContext](#).

Example Usar o shell do Spark para contar as ocorrências de uma string em um arquivo armazenado no Amazon S3

Este exemplo usa `sc` para ler um arquivo de texto armazenado no 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")
```

O Spark cria a [estrutura `textFile` de dados](#) associada. Em seguida, o exemplo conta o número de linhas no arquivo de log com a string "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
```

Example Usar o shell do Spark baseado em Python para contar as ocorrências de uma string em um arquivo armazenado no Amazon S3

O Spark também inclui um shell baseado em Python, `pyspark`, que você pode usar para gerar protótipos de programas Spark escritos em Python. Assim como `comspark-shell`, invoque `pyspark` no nó primário; ele também tem o mesmo [SparkContext](#) objeto.

```
>>> 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")
```

O Spark cria a [estrutura `textFile` de dados](#) associada. Em seguida, o exemplo conta o número de linhas no arquivo de log com a string "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
```

## Use o Amazon SageMaker Spark para aprendizado de máquina

Ao usar a Amazon EMR versão 5.11.0 e posterior, o `aws-sagemaker-spark-sdk` componente é instalado junto com o Spark. [Esse componente instala o Amazon SageMaker Spark e as dependências associadas para a integração do Spark com a Amazon SageMaker](#) Você pode usar o Amazon SageMaker Spark para criar pipelines de aprendizado de máquina (ML) do Spark usando os estágios da Amazon SageMaker Para obter mais informações, consulte o [Amazon SageMaker Spark README](#) sobre GitHub e como [usar o Apache Spark com a Amazon no Amazon SageMaker Developer Guide](#). SageMaker

## Escrever uma aplicação do Spark

Aplicativos [Spark](#) podem ser escritos em Scala, Java ou Python. Existem vários exemplos de aplicações do Spark no tópico [Spark examples](#) da documentação do Apache Spark. O exemplo "Estimating Pi" é mostrado abaixo em três aplicativos com suporte nativo. Você também pode ver exemplos completos em `$SPARK_HOME/examples` e em [GitHub](#). Para obter mais informações sobre como criar JARs para o Spark, consulte o tópico [Início rápido](#) na documentação do Apache Spark.

### Scala

Para evitar problemas de compatibilidade com o Scala, sugerimos que você use as dependências do Spark para a versão correta do Scala ao compilar um aplicativo Spark para um cluster da Amazon EMR A versão do Scala que você deve usar depende da versão do Spark instalada no cluster. Por exemplo, a EMR versão 5.30.1 da Amazon usa o Spark 2.4.5, que é construído com o Scala

2.11. Se seu cluster usa a EMR versão 5.30.1 da Amazon, use as dependências do Spark para o Scala 2.11. Para obter mais informações sobre as versões do Scala usadas pelo Spark, consulte a [documentação do 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)
```

## Melhorar a performance do Spark com Amazon S3

A Amazon EMR oferece recursos para ajudar a otimizar o desempenho ao usar o Spark para consultar, ler e gravar dados salvos no Amazon S3.

[O S3 Select](#) pode melhorar o desempenho de consultas CSV e JSON arquivos em alguns aplicativos “transferindo” o processamento para o Amazon S3.

O committer EMRFS otimizado para S3 é uma alternativa à [OutputCommitter](#) classe, que usa o recurso de uploads de várias partes EMRFS para melhorar o desempenho ao gravar arquivos Parquet no Amazon S3 usando Spark e conjuntos de dados. SQL DataFrames

### Tópicos

- [Usar o S3 Select com Spark para melhorar a performance das consultas](#)
- [Use o committer EMRFS otimizado para S3](#)
- [Use o protocolo de confirmação EMRFS otimizado para S3](#)
- [Tente novamente as solicitações do Amazon S3 com EMRFS](#)

## Usar o S3 Select com Spark para melhorar a performance das consultas

Com a EMR versão 5.17.0 da Amazon e versões posteriores, você pode usar o [S3 Select](#) com o Spark na Amazon. O S3 Select possibilita que as aplicações recuperem apenas um subconjunto dos dados de um objeto. Para a AmazonEMR, o trabalho computacional de filtrar grandes conjuntos de dados para processamento é “transferido” do cluster para o Amazon S3, o que pode melhorar o desempenho em alguns aplicativos e reduzir a quantidade de dados transferidos entre a Amazon e o EMR Amazon S3.

O S3 Select é compatível com JSON arquivos CSV `s3selectCSV` e `s3selectJSON` valores para especificar o formato dos dados. Para ter mais informações e exemplos, consulte [Especificar o S3 Select no seu código](#).

## O S3 Select é adequado para minha aplicação?

Recomendamos que você avalie seus aplicativos com e sem o S3 Seleção para ver se o uso pode ser adequado para o aplicativo.

Use as seguintes diretrizes para determinar se o seu aplicativo é adequado para o uso do S3 Select:

- Sua consulta filtra mais de metade do conjunto de dados original.
- Sua conexão de rede entre o Amazon S3 e o EMR cluster da Amazon tem boa velocidade de transferência e largura de banda disponível. O Amazon S3 não compacta HTTP respostas, então é provável que o tamanho da resposta aumente para arquivos de entrada compactados.

## Considerações e limitações

- A criptografia do lado do servidor do Amazon S3 com chaves de criptografia fornecidas pelo cliente (SSE-C) e a criptografia do lado do cliente não são suportadas.
- A propriedade `AllowQuotedRecordDelimiters` não é compatível. Se essa propriedade for especificada, a consulta falhará.
- Somente JSON arquivos CSV e arquivos no formato UTF -8 são suportados. Não CSVs há suporte para várias linhas.
- Somente arquivos descompactados ou gzip são compatíveis.
- O Spark CSV e JSON opções `comonanValue`, `positiveInfnegativeInf`, e opções relacionadas a registros corrompidos (por exemplo, modo `failfast` e `dropmalformed`) não são suportados.
- O uso de vírgulas (,) em casas decimais não é compatível. Por exemplo, `10,000` não é compatível, mas `10000` é.
- Caracteres de comentário na última linha não são compatíveis.
- Linhas vazias no final de um arquivo não são processadas.
- Os seguintes filtros não são enviados para o Amazon S3:
  - Funções agregadas, como `COUNT()` e `SUM()`.
  - Filtros que `CAST()` um atributo. Por exemplo, `CAST(stringColumn as INT) = 1`.

- Filtros com um atributo que é um objeto ou complexo. Por exemplo, `intArray[1] = 1`, `objectColumn.objectNumber = 1`.
- Filtros para os quais o valor não é um valor literal. Por exemplo, `intColumn1 = intColumn2`
- Somente [tipos de dados compatíveis com o S3 Select](#) são compatíveis com as limitações documentadas.

## Especificar o S3 Select no seu código

Os exemplos a seguir demonstram como especificar o S3 Select para CSV usar Scala, SQL R e PySpark. Você pode usar o S3 Select for JSON da mesma forma. Para obter uma lista de opções, os valores padrão e limitações, consulte [Opções](#).

### 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")
```

## Opções

As seguintes opções estão disponíveis ao usar `s3selectCSV` e `s3selectJSON`. Se não for especificado, os valores padrão serão usados.

## Opções com S3select CSV

| Opção                    | Padrão  | Uso                                                                                                                                                                                |
|--------------------------|---------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>compression</code> | "none"  | Indica se a compactação é usada. "gzip" é a única configuração compatível além de "none".                                                                                          |
| <code>delimiter</code>   | ","     | Especifica o delimitador de campo.                                                                                                                                                 |
| <code>quote</code>       | '\"'    | Especifica o caractere de aspas. A especificação de uma string vazia não é suportada e resulta em um erro malformatadoXML.                                                         |
| <code>escape</code>      | '\\'    | Especifica o caractere de escape.                                                                                                                                                  |
| <code>header</code>      | "false" | "false" especifica que não há cabeçalho. "true" especifica que o cabeçalho está na primeira linha. Somente cabeçalhos na primeira linha são suportados e linhas vazias antes de um |



| Opção      | Padrão | Uso                                                                                                                                                            |
|------------|--------|----------------------------------------------------------------------------------------------------------------------------------------------------------------|
|            |        | cabeçalho não são compatíveis.                                                                                                                                 |
| comentário | "#"    | Especifica o caractere de comentário. O indicador de comentários não pode ser desativado. Em outras palavras, um valor de <code>\u0000</code> não é suportado. |
| nullValue  | ""     |                                                                                                                                                                |

### Opções com S3select JSON

| Opção       | Padrão  | Uso                                                                                                                                                                                                                                                                                                   |
|-------------|---------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| compression | "none"  | Indica se a compactação é usada. "gzip" é a única configuração compatível além de "none".                                                                                                                                                                                                             |
| multiline   | "falso" | "false" especifica que o JSON está no LINES formato S3 Select, o que significa que cada linha nos dados de entrada contém um único JSON objeto. "true" especifica que o JSON está no DOCUMENT formato S3 Select, o que significa que um JSON objeto pode abranger várias linhas nos dados de entrada. |

## Use o committer EMRFS otimizado para S3

O committer EMRFS otimizado para S3 é uma [OutputCommitter](#) implementação alternativa que é otimizada para gravar arquivos no Amazon S3 durante o uso. O committer EMRFS otimizado para S3 melhora o desempenho do aplicativo ao evitar operações de lista e renomeação realizadas no Amazon S3 durante as fases de confirmação de tarefas e trabalhos. O committer está disponível com a Amazon EMR versão 5.19.0 e posterior e está habilitado por padrão com o Amazon EMR 5.20.0 e versões posteriores. O committer é usado para trabalhos do Spark que usam Spark ou conjuntos de SQL dados DataFrames. A partir do Amazon EMR 6.4.0, esse committer pode ser usado para todos os formatos comuns ORC, incluindo parquet e formatos baseados em texto (incluindo e). CSV JSON Para versões anteriores ao Amazon EMR 6.4.0, somente o formato Parquet é suportado. Há circunstâncias em que o committer não é usado. Para obter mais informações, consulte [Requisitos para o committer otimizado para EMRFS S3](#).

### Tópicos

- [Requisitos para o committer otimizado para EMRFS S3](#)
- [O committer EMRFS otimizado para S3 e os uploads em várias partes](#)
- [Considerações sobre ajuste de tarefas](#)
- [Habilite o committer EMRFS otimizado para S3 para o Amazon 5.19.0 EMR](#)

### Requisitos para o committer otimizado para EMRFS S3


O committer EMRFS otimizado para S3 é usado quando as seguintes condições são atendidas:

- Você executa trabalhos do Spark que usam o Spark SQL ou conjuntos de dados para gravar arquivos no Amazon S3. DataFrames A partir do Amazon EMR 6.4.0, esse committer pode ser usado para todos os formatos comuns ORC, incluindo parquet e formatos baseados em texto (incluindo e). CSV JSON Para versões anteriores ao Amazon EMR 6.4.0, somente o formato Parquet é suportado.
- Os uploads de várias partes estão habilitados na Amazon. EMR Esse é o padrão. Para obter mais informações, consulte [O committer EMRFS otimizado para S3 e os uploads em várias partes](#).
- O suporte a formato de arquivo integrado do Spark é usado. O suporte a formato de arquivo integrado é usado nas seguintes circunstâncias:
  - Para tabelas do Hive Metastore, quando `spark.sql.hive.convertMetastoreParquet` está definido como `true` para tabelas Parquet ou definido como `true` para tabelas

`spark.sql.hive.convertMetastoreOrc` Orc com Amazon EMR 6.4.0 ou superior. Essas são as configurações padrão.

- Quando os trabalhos são gravados em fontes de dados ou tabelas do Parquet, por exemplo, a tabela de destino é criada com a cláusula `USING parquet`.
- Quando os trabalhos gravam em tabelas Parquet de metastore do Hive não particionadas. O suporte ao Parquet incorporado do Spark oferece suporte a tabelas particionadas do Hive, o que é uma limitação conhecida. Para obter mais informações, consulte [Conversão da tabela Parquet do Hive metastore](#) no Apache SQL Spark e no Guia de conjuntos de dados. DataFrames
- As operações de trabalhos do Spark que gravam em um local de partição padrão, como `${table_location}/k1=v1/k2=v2/`, usam o confirmador. O confirmador não será usado se uma operação de trabalho gravar em um local de partição personalizado, por exemplo, se o local de uma partição personalizado for definido usando o comando `ALTER TABLE SQL`.
- Os valores a seguir para o Spark devem ser usados:
  - A propriedade `spark.sql.parquet.fs.optimized.committer.optimization-enabled` deve ser definida como `true`. Essa é a configuração padrão com o Amazon EMR 5.20.0 e versões posteriores. Com o Amazon EMR 5.19.0, o valor padrão é `false` Para obter informações sobre como configurar esse valor, consulte [Habilite o committer EMRFS otimizado para S3 para o Amazon 5.19.0 EMR](#).
  - Se estiver gravando em tabelas de metastore Hive não particionadas, somente os formatos de arquivo Parquet e Orc são suportados. `spark.sql.hive.convertMetastoreParquet` deve ser definido como `true` se estiver gravando em tabelas de metastore não particionadas do Parquet Hive. `spark.sql.hive.convertMetastoreOrc` deve ser definido como `true` se estiver gravando em tabelas não particionadas do metastore Orc Hive. Essas são as configurações padrão.
  - `spark.sql.parquet.output.committer.class` deve ser definido como `com.amazon.emr.committer.EmrOptimizedSparkSqlParquetOutputCommitter`. Essa é a configuração padrão.
  - `spark.sql.sources.commitProtocolClass` deve ser definido como `org.apache.spark.sql.execution.datasources.SQLEmrOptimizedCommitProtocol` ou `org.apache.spark.sql.execution.datasources.SQLHadoopMapReduceCommitProtocol` `org.apache.spark.sql.execution.datasources.SQLEmrOptimizedCommitProtocol` é a configuração padrão para a série EMR 5.x da Amazon versão 5.30.0 e superior e para a série EMR 6.x da Amazon versão 6.2.0 e superior. `org.apache.spark.sql.execution.datasources.SQLHadoopMapReduceCommitProtocol` é a configuração padrão para as EMR versões anteriores da Amazon.

- Se os trabalhos do Spark substituírem os conjuntos de dados Parquet por colunas de partição dinâmica, as opções de gravação `partitionOverwriteMode` e `spark.sql.sources.partitionOverwriteMode` deverão ser definidas como `static`. Essa é a configuração padrão.

 Note

A opção de gravação `partitionOverwriteMode` foi introduzida no Spark 2.4.0. Para a versão 2.3.2 do Spark, incluída na EMR versão 5.19.0 da Amazon, defina a propriedade `spark.sql.sources.partitionOverwriteMode`

Ocasões em que o committer EMRFS otimizado para S3 não é usado

Geralmente, o committer EMRFS otimizado para S3 não é usado nas seguintes situações.

| Situação                                        | Por que o confirmador não é usado                                       |
|-------------------------------------------------|-------------------------------------------------------------------------|
| Quando você escreve para HDFS                   | O committer só suporta a gravação no Amazon EMRFS S3 usando.            |
| Quando você usa o sistema de arquivos S3A       | O comitê só apóia EMRFS.                                                |
| Quando você usa MapReduce nosso Spark's RDD API | O committer só suporta o uso do Spark ou do SQL DataFrame Dataset. APIs |

Os exemplos de Scala a seguir demonstram algumas situações adicionais que impedem que o committer EMRFS otimizado para S3 seja usado no todo (o primeiro exemplo) e em parte (o segundo exemplo).

#### Example – Modo de substituição de partição dinâmica

O exemplo de Scala a seguir instrui o Spark a usar um algoritmo de confirmação diferente, o que impede totalmente o uso do committer otimizado para EMRFS S3. O código define a propriedade `partitionOverwriteMode` como `dynamic` para substituir somente as partições nas quais você está gravando dados. Em seguida, as colunas de partição dinâmica são especificadas por `partitionBy` e o modo de gravação é definido como `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")
```

Você deve definir todas as três configurações para evitar o uso do committer EMRFS otimizado para S3. Quando você faz isso, o Spark executa um algoritmo de confirmação diferente, especificado no protocolo de confirmação do Spark. Para versões do Amazon EMR 5.x anteriores à 5.30.0 e para versões do Amazon EMR 6.x anteriores à 6.2.0, o protocolo de confirmação usa o diretório de teste do Spark, que é um diretório temporário criado sob o local de saída que começa com `.spark-staging`. O algoritmo renomeia sequencialmente diretórios de partição, o que pode afetar negativamente a performance. Para obter mais informações sobre as EMR versões 5.30.0 e posteriores e 6.2.0 e posteriores da Amazon, consulte [Use o protocolo de confirmação EMRFS otimizado para S3](#)

O algoritmo no Spark 2.4.0 segue estas etapas:

1. As tentativas de tarefa gravam a saída nos diretórios de partição do diretório de teste do Spark, por exemplo, `${outputLocation}/spark-staging-${jobID}/k1=v1/k2=v2/`.
2. Para cada partição gravada, a tentativa de tarefa acompanha os caminhos de partição relativos, por exemplo, `k1=v1/k2=v2`.
3. Quando uma tarefa é concluída com êxito, ela fornece o driver com todos os caminhos de partição relativos que ela controlou.
4. Depois que todas as tarefas forem concluídas, a fase de confirmação do trabalho coletará todos os diretórios da partição que as tentativas de tarefas bem-sucedidas gravaram no diretório de preparação do Spark. O Spark renomeia sequencialmente cada um desses diretórios para o local de saída final usando as operações para renomear a árvore de diretórios.
5. O diretório de preparação é excluído antes de a fase de confirmação de trabalho ser concluída.

#### Example – Local de partição personalizado

Neste exemplo, o código Scala insere em duas partições. Uma partição tem um local de partição personalizado. A outra partição usa o local de partição padrão. O committer EMRFS otimizado para

S3 é usado somente para gravar a saída da tarefa na partição que usa a localização padrão da partição.

```
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)
```

O código Scala cria os seguintes objetos do Amazon S3:

```
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_$folder$
```

Ao gravar em partições em locais personalizados, o Spark usa um algoritmo de confirmação semelhante ao exemplo anterior, que é descrito abaixo. Como no exemplo anterior, o algoritmo resulta em renomeações sequenciais, o que pode afetar negativamente a performance.

1. Ao gravar a saída em uma partição em um local personalizado, as tarefas gravam em um arquivo no diretório de preparação do Spark, que é criado no local de saída final. O nome do arquivo inclui um aleatório UUID para proteger contra colisões de arquivos. A tentativa de tarefa controla cada arquivo junto com o caminho de saída final desejado.
2. Quando uma tarefa é concluída com êxito, ela fornece o driver com os arquivos e os caminhos desejados de saída final.
3. Depois que todas as tarefas forem concluídas, a fase de confirmação do trabalho renomeará sequencialmente todos os arquivos que foram gravados para partições em locais personalizados em seus caminhos de saída final.
4. O diretório de preparação é excluído antes de a fase de confirmação de trabalho ser concluída.

## O committer EMRFS otimizado para S3 e os uploads em várias partes

Para usar o committer EMRFS otimizado para S3, você deve habilitar uploads de várias partes para a Amazon. EMR Multipart uploads são habilitados por padrão. Você pode habilitá-los novamente, se necessário. Para obter mais informações, consulte [Configurar o upload de várias partes para o Amazon S3](#) no EMRAmazon Management Guide.

O committer EMRFS otimizado para S3 usa as características de transação de uploads de várias partes para garantir que os arquivos gravados por tentativas de tarefa só apareçam no local de saída do trabalho após a confirmação da tarefa. Ao usar uploads de várias partes dessa forma, o committer melhora o desempenho da confirmação da tarefa em relação à versão 2 do FileOutputCommitter algoritmo padrão. Ao usar o committer EMRFS otimizado para S3, há algumas diferenças importantes em relação ao comportamento tradicional de upload em várias partes a serem consideradas:

- Os multipart uploads são sempre executados, independentemente do tamanho do arquivo. Isso difere do comportamento padrão de EMRFS, em que a `fs.s3n.multipart.uploads.split.size` propriedade controla o tamanho do arquivo no qual os uploads de várias partes são acionados.
- Os multipart uploads são deixados incompletos por um período mais longo até que a tarefa seja confirmada ou cancelada. Isso difere do comportamento padrão de EMRFS quando um upload de várias partes é concluído quando uma tarefa termina de gravar um determinado arquivo.

Por causa dessas diferenças, se um executor do Spark JVM falhar ou for interrompido enquanto as tarefas estão sendo executadas e gravando dados no Amazon S3, é mais provável que os uploads

incompletos de várias partes sejam deixados para trás. Por esse motivo, ao usar o committer EMRFS otimizado para S3, certifique-se de seguir as melhores práticas para gerenciar uploads de várias partes com falha. Para obter mais informações, consulte [Melhores práticas](#) para trabalhar com buckets do Amazon S3 no Guia de gerenciamento da Amazon EMR.

## Considerações sobre ajuste de tarefas

O committer EMRFS otimizado para S3 consome uma pequena quantidade de memória para cada arquivo gravado por uma tentativa de tarefa até que a tarefa seja confirmada ou abortada. Na maioria dos trabalhos, a quantidade de memória consumida é insignificante. Para trabalhos que têm tarefas de longa execução que gravam um grande número de arquivos, a memória que o committer consome pode ser perceptível e exigem ajustes na memória alocada para executores do Spark. É possível ajustar a memória do executor usando a propriedade `spark.executor.memory`. Como diretriz, uma única tarefa gravando 100.000 arquivos normalmente exigiria 100 MB adicionais de memória. Para obter mais informações, consulte [Propriedades da aplicação](#) na documentação de configuração do Apache Spark.

## Habilite o committer EMRFS otimizado para S3 para o Amazon 5.19.0 EMR

Se você estiver usando o Amazon EMR 5.19.0, poderá definir manualmente a `spark.sql.parquet.fs.optimized.committer.optimization-enabled` propriedade para `true` quando criar um cluster ou de dentro do Spark se estiver usando a Amazon EMR

Habilitando o committer EMRFS otimizado para S3 ao criar um cluster

Use a classificação de configuração `spark-defaults` para definir a propriedade `spark.sql.parquet.fs.optimized.committer.optimization-enabled` como `true`. Para obter mais informações, consulte [Configurar aplicações](#).

Habilitando o committer EMRFS otimizado para S3 do Spark

Você pode definir `spark.sql.parquet.fs.optimized.committer.optimization-enabled` como `true` codificando-o em um `SparkConf`, transmitindo-o como um parâmetro `--conf` no shell Spark ou nas ferramentas `spark-submit` e `spark-sql` ou em `conf/spark-defaults.conf`. Para obter mais informações, consulte [Configuração do Spark](#) na documentação do Apache Spark.

O exemplo a seguir mostra como habilitar o committer ao executar um comando `spark-sql`.

```
spark-sql \
```



```
--conf spark.sql.parquet.fs.optimized.committer.optimization-enabled=true \
-e "INSERT OVERWRITE TABLE target_table SELECT * FROM source_table;"
```

## Use o protocolo de confirmação EMRFS otimizado para S3

O protocolo de confirmação EMRFS otimizado para S3 é uma [FileCommitProtocol](#) implementação alternativa que é otimizada para gravar arquivos com a substituição de partição dinâmica do Spark no Amazon S3 durante o uso. O protocolo EMRFS melhora a performance da aplicação ao evitar operações de renomeação no Amazon S3 durante a fase de confirmação do trabalho de substituição de partição dinâmica do Spark.

Observe que [Use o committer EMRFS otimizado para S3](#) também melhora o a performance ao evitar operações de renomeação. No entanto, ele não funciona para casos de substituição de partição dinâmica, embora as melhorias no protocolo de confirmação sejam direcionados apenas a casos de substituição de partição dinâmica.

O protocolo de confirmação está disponível com a EMR versão 5.30.0 e posterior da Amazon e 6.2.0 e posterior e está habilitado por padrão. A Amazon EMR adicionou uma melhoria de paralelismo a partir da versão 5.31.0. O protocolo é usado para trabalhos do Spark que usam Spark ou conjuntos SQL de DataFrames dados. Há circunstâncias em que o protocolo de confirmação não é usado. Para obter mais informações, consulte [Requisitos para o protocolo de EMRFS confirmação otimizado para S3](#).

### Tópicos

- [Requisitos para o protocolo de EMRFS confirmação otimizado para S3](#)
- [O protocolo de confirmação EMRFS otimizado para S3 e uploads em várias partes](#)
- [Considerações sobre ajuste de tarefas](#)

## Requisitos para o protocolo de EMRFS confirmação otimizado para S3

O protocolo de confirmação EMRFS otimizado para S3 é usado quando as seguintes condições são atendidas:

- Você executa trabalhos do Spark que usam o Spark SQL ou conjuntos de dados para substituir tabelas particionadas. DataFrames
- Você executa trabalhos do Spark cujo modo de substituição de partição é `dynamic`.

- Os uploads de várias partes estão habilitados na Amazon. EMR Esse é o padrão. Para obter mais informações, consulte [O protocolo de confirmação EMRFS otimizado para S3 e uploads em várias partes](#).
- O cache do sistema de arquivos para EMRFS está habilitado. Esse é o padrão. Verifique se a configuração `fs.s3.impl.disable.cache` está definida como `false`.
- O suporte integrado de fonte de dados do Spark é usado. O suporte integrado ao Parquet é usado nas seguintes circunstâncias:
  - Quando os trabalhos gravam em fontes de dados ou tabelas integradas.
  - Quando os trabalhos gravam em tabelas do Parquet do metastore do Hive. Isso acontece quando `spark.sql.hive.convertInsertingPartitionedTable` e `spark.sql.hive.convertMetastoreParquet` são definidos como verdadeiros. Essas são as configurações padrão.
  - Quando os trabalhos são gravados na tabela do ORC metastore do Hive. Isso acontece quando `spark.sql.hive.convertInsertingPartitionedTable` e `spark.sql.hive.convertMetastoreOrc` são definidos como `true`. Essas são as configurações padrão.
- As operações de trabalhos do Spark que gravam em um local de partição padrão, por exemplo, `${table_location}/k1=v1/k2=v2/`, usam o confirmador. O protocolo não será usado se uma operação de trabalho gravar em um local de partição personalizado, por exemplo, se o local de uma partição personalizado for definido usando o comando `ALTER TABLE SQL`.
- Os valores a seguir para o Spark devem ser usados:
  - `spark.sql.sources.commitProtocolClass` deve ser definido como `org.apache.spark.sql.execution.datasources.SQLEmrOptimizedCommitProtocol`. Essa é a configuração padrão para as EMR versões 5.30.0 e superiores da Amazon e 6.2.0 e superiores.
  - A opção de gravação `partitionOverwriteMode` ou `spark.sql.sources.partitionOverwriteMode` deve ser definida como `dynamic`. A configuração padrão é `static`.

**Note**

A opção de gravação `partitionOverwriteMode` foi introduzida no Spark 2.4.0. Para a versão 2.3.2 do Spark, incluída na EMR versão 5.19.0 da Amazon, defina a propriedade `spark.sql.sources.partitionOverwriteMode`

- Se os trabalhos do Spark substituírem a tabela do Parquet do metastore do Hive, `spark.sql.hive.convertMetastoreParquet`, `spark.sql.hive.convertInsertingPartitionedTable` e `spark.sql.hive.convertMetastore.partitionOverwriteMode` deverão ser configurados como `true`. Existem as configurações padrão.
- Se os trabalhos do Spark forem substituídos pela ORC tabela do metastore do Hive,, `spark.sql.hive.convertMetastoreOrc` `spark.sql.hive.convertInsertingPartitionedTable` e `spark.sql.hive.convertMetastore.partitionOverwriteMode` devem ser definidos como `true`. Existem as configurações padrão.

### Example – Modo de substituição de partição dinâmica

Neste exemplo do Scala, a otimização é acionada. Primeiro, você define a propriedade `partitionOverwriteMode` como `dynamic`. Isso só substitui as partições nas quais você está gravando dados. Em seguida, você especifica as colunas de partição dinâmica com `partitionBy` e define o modo de gravação como `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://"
```

### Quando o protocolo de confirmação EMRFS otimizado para S3 não é usado

Geralmente, o protocolo de confirmação EMRFS otimizado para S3 funciona da mesma forma que o protocolo de SQL confirmação Spark padrão de código aberto, `org.apache.spark.sql.execution.datasources.SQLHadoopMapReduceCommitProtocol`. A otimização não ocorrerá nas situações a seguir.

|                                                           |                                                                                                                                                                            |
|-----------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Situação                                                  | Por que o protocolo de confirmação não é usado                                                                                                                             |
| Quando você escreve para HDFS                             | O protocolo de confirmação só suporta a gravação no Amazon S3 usando EMRFS                                                                                                 |
| Quando você usa o sistema de arquivos S3A                 | O protocolo de confirmação só oferece suporte EMRFS.                                                                                                                       |
| Quando você usa MapReduce nosso Spark's RDD API           | O protocolo de confirmação só oferece suporte ao uso do Spark SQL ou do APIs Dataset, DataFrame                                                                            |
| Quando a substituição da partição dinâmica não é acionada | O protocolo de confirmação só otimiza os casos de substituição de partição dinâmica. Para outros casos, consulte <a href="#">Use o committer EMRFS otimizado para S3</a> . |

Os exemplos de Scala a seguir demonstram algumas situações adicionais às quais o protocolo de confirmação EMRFS otimizado para S3 delega. `SQLHadoopMapReduceCommitProtocol`

Example – Modo de substituição de partição dinâmica com local de partição personalizado

Neste exemplo, os programas Scala sobrescrevem duas partições no modo de substituição dinâmica de partição. Uma partição tem um local de partição personalizado. A outra partição usa o local de partição padrão. O protocolo de confirmação EMRFS otimizado para S3 só melhora a partição que usa a localização padrão da partição.

```
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")

```

O código Scala cria os seguintes objetos do Amazon S3:

```

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

Gravar em locais de partição personalizados em versões anteriores do Spark pode resultar em perda de dados. Neste exemplo, a partição dt='2019-01-28' seria perdida. Para obter mais detalhes, consulte [SPARK-35106](#). Isso foi corrigido na EMR versão 5.33.0 e posterior da Amazon, excluindo 6.0.x e 6.1.x.

Ao gravar em partições em locais personalizados, o Spark usa um algoritmo de confirmação semelhante ao exemplo anterior, que é descrito abaixo. Como no exemplo anterior, o algoritmo resulta em renomeações sequenciais, o que pode afetar negativamente a performance.

O algoritmo no Spark 2.4.0 segue estas etapas:

1. Ao gravar a saída em uma partição em um local personalizado, as tarefas gravam em um arquivo no diretório de preparação do Spark, que é criado no local de saída final. O nome do arquivo inclui um aleatório UUID para proteger contra colisões de arquivos. A tentativa de tarefa controla cada arquivo junto com o caminho de saída final desejado.
2. Quando uma tarefa é concluída com êxito, ela fornece o driver com os arquivos e os caminhos desejados de saída final.
3. Depois que todas as tarefas forem concluídas, a fase de confirmação do trabalho renomeará sequencialmente todos os arquivos que foram gravados para partições em locais personalizados em seus caminhos de saída final.
4. O diretório de preparação é excluído antes de a fase de confirmação de trabalho ser concluída.

## O protocolo de confirmação EMRFS otimizado para S3 e uploads em várias partes

Para usar a otimização para substituição dinâmica de partições no protocolo de confirmação EMRFS otimizado para S3, os uploads de várias partes devem estar habilitados na Amazon. EMR Multipart uploads são habilitados por padrão. Você pode habilitá-los novamente, se necessário. Para obter mais informações, consulte [Configurar o upload de várias partes para o Amazon S3](#) no EMRAmazon Management Guide.

Durante a substituição dinâmica da partição, o protocolo de confirmação EMRFS otimizado para S3 usa as características de transação dos carregamentos de várias partes para garantir que os arquivos gravados por tentativas de tarefa apareçam apenas no local de saída do trabalho após a confirmação do trabalho. Ao usar carregamentos multipart dessa maneira, o protocolo de confirmação melhora a performance de confirmação de trabalhos em relação ao padrão `SQLHadoopMapReduceCommitProtocol`. Ao usar o protocolo de confirmação EMRFS otimizado para S3, há algumas diferenças importantes em relação ao comportamento tradicional de upload em várias partes a serem consideradas:

- Os multipart uploads são sempre executados, independentemente do tamanho do arquivo. Isso difere do comportamento padrão de EMRFS, em que a `fs.s3n.multipart.uploads.split.size` propriedade controla o tamanho do arquivo no qual os uploads de várias partes são acionados.
- Os multipart uploads são deixados incompletos por um período mais longo até que a tarefa seja confirmada ou cancelada. Isso difere do comportamento padrão de EMRFS quando um upload de várias partes é concluído quando uma tarefa termina de gravar um determinado arquivo.

Devido a essas diferenças, se um executor do Spark JVM falhar ou for encerrado enquanto as tarefas estão sendo executadas e gravando dados no Amazon S3, ou se um driver do Spark JVM falhar ou for encerrado durante a execução de um trabalho, é mais provável que os uploads incompletos de várias partes sejam deixados para trás. Por esse motivo, ao usar o protocolo de confirmação EMRFS otimizado para S3, certifique-se de seguir as melhores práticas para gerenciar uploads de várias partes com falha. Para obter mais informações, consulte [Melhores práticas](#) para trabalhar com buckets do Amazon S3 no Guia de gerenciamento da Amazon EMR.

## Considerações sobre ajuste de tarefas

Nos executores do Spark, o protocolo de confirmação EMRFS otimizado para S3 consome uma pequena quantidade de memória para cada arquivo gravado por uma tentativa de tarefa até que a tarefa seja confirmada ou abortada. Na maioria dos trabalhos, a quantidade de memória consumida é insignificante.

Nos drivers do Spark, o protocolo de confirmação EMRFS otimizado para S3 requer memória para armazenar informações de metadados de cada arquivo confirmado até que o trabalho seja confirmado ou abortado. Na maioria dos trabalhos, a configuração padrão de memória do driver do Spark é insignificante.

Para trabalhos que têm tarefas de execução prolongada que gravam um grande número de arquivos, a memória que o confirmador consome pode ser perceptível e precisar de ajustes na memória alocada para executores do Spark. Você pode ajustar a memória usando a `spark.driver.memory` propriedade dos drivers do Spark e a `spark.executor.memory` propriedade dos executores do Spark. Como diretriz, uma única tarefa gravando 100.000 arquivos normalmente exigiria 100 MB adicionais de memória. Para obter mais informações, consulte [Propriedades da aplicação](#) na documentação de configuração do Apache Spark.

## Tente novamente as solicitações do Amazon S3 com EMRFS

Este tópico fornece informações sobre as estratégias de repetição que você pode usar ao fazer solicitações para o Amazon EMRFS S3 com. Quando a taxa de solicitação aumenta, o S3 tenta escalar para suportar a nova taxa. Durante esse processo, o S3 pode controlar a utilização das solicitações e retornar um erro 503 `Slow Down`. Para melhorar a taxa de sucesso das solicitações do S3, você pode ajustar sua estratégia de novas tentativas configurando propriedades na sua configuração `emrfs-site`.

Você pode ajustar sua estratégia de novas tentativas das maneiras a seguir.

- Aumente o limite máximo de novas tentativas para a estratégia padrão de novas tentativas de recuo exponencial.
- Ative e configure a estratégia de nova tentativa de aumento aditivo/diminuição multiplicativa (). AIMD AIMD é compatível com as EMR versões 6.4.0 e posteriores da Amazon.

## Usar a estratégia padrão de recuo exponencial

Por padrão, EMRFS usa uma estratégia de recuo exponencial para repetir as solicitações do Amazon S3. O limite padrão EMRFS de repetição é 15. Para evitar um erro 503 Slow Down do S3, você pode aumentar o limite de novas tentativas ao criar um novo cluster, em um cluster em execução ou no runtime da aplicação.

Para aumentar o limite de novas tentativas, você deve alterar o valor de `fs.s3.maxRetries` na sua configuração `emrfs-site`. O exemplo de configuração a seguir define `fs.s3.maxRetries` como um valor personalizado de 30.

```
[
 {
 "Classification": "emrfs-site",
 "Properties": {
 "fs.s3.maxRetries": "30"
 }
 }
]
```

Para obter mais informações sobre como trabalhar com objetos de configuração, consulte [Configurar aplicações](#).

## Use a estratégia AIMD de nova tentativa

Com a EMR versão 6.4.0 e posterior da Amazon, EMRFS oferece suporte a uma estratégia alternativa de repetição com base em um modelo de aumento aditivo/diminuição multiplicativa (). AIMD A estratégia de AIMD repetição é especialmente útil quando você trabalha com grandes EMR clusters da Amazon.

AIMD calcula uma taxa de solicitação personalizada usando dados sobre solicitações recentes bem-sucedidas. Essa estratégia diminui o número de solicitações submetidas a controle de utilização e o total de tentativas necessárias por solicitação.



Para habilitar a estratégia de AIMD repetição, você deve definir a `fs.s3.aimd.enabled` propriedade como `true` em sua `emrfs-site` configuração, como no exemplo a seguir.

```
[
 {
 "Classification": "emrfs-site",
 "Properties": {
 "fs.s3.aimd.enabled": "true"
 }
 }
]
```

Para obter mais informações sobre como trabalhar com objetos de configuração, consulte [Configurar aplicações](#).

## Configurações avançadas AIMD de nova tentativa

Você pode configurar as propriedades listadas na tabela a seguir para refinar o comportamento de repetição ao usar a estratégia de AIMD repetição. Para a maioria dos casos de uso, recomendamos que você use os valores padrão.

### Propriedades avançadas AIMD da estratégia de repetição

| Propriedade                               | Valor padrão | Descrição                                                                                                                                                          |
|-------------------------------------------|--------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>fs.s3.aimd.increaseIncrement</code> | 0.1          | Controla a rapidez com que a taxa de solicitações aumenta quando solicitações consecutivas são bem-sucedidas.                                                      |
| <code>fs.s3.aimd.reductionFactor</code>   | 2            | Controla a rapidez com que a taxa de solicitação diminui quando o Amazon S3 retorna uma resposta 503. O fator padrão de 2 reduz a taxa de solicitação pela metade. |
| <code>fs.s3.aimd.minRate</code>           | 0.1          | Define o limite inferior da taxa de solicitações quando as solicitações sofrem controle                                                                            |

| Propriedade                          | Valor padrão | Descrição                                                                                                                                                                                                                                                                                                                     |
|--------------------------------------|--------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                      |              | de utilização sustentado pelo S3.                                                                                                                                                                                                                                                                                             |
| <code>fs.s3.aimd.initialRate</code>  | 5500         | Define a taxa de solicitação inicial, que sofre alterações de acordo com os valores que você especifica para <code>fs.s3.aimd.increaseIncrement</code> e <code>fs.s3.aimd.reductionFactor</code> .<br><br>A taxa inicial também é usada para GET solicitações e escalada proporcionalmente (3500/5500) para solicitações. PUT |
| <code>fs.s3.aimd.adjustWindow</code> | 2            | Controla a frequência com que a taxa de solicitação é ajustada, medida em número de respostas.                                                                                                                                                                                                                                |
| <code>fs.s3.aimd.maxAttempts</code>  | 100          | Define o número máximo de tentativas para testar uma solicitação.                                                                                                                                                                                                                                                             |

## Adicionar uma etapa do Spark

Você pode usar EMR as etapas da Amazon para enviar trabalhos para a estrutura do Spark instalada em um EMR cluster. Para obter mais informações, consulte [as etapas](#) no Guia EMR de gerenciamento da Amazon. No consoleCLI, você faz isso usando uma etapa do aplicativo Spark, que executa o `spark-submit` script como uma etapa em seu nome. Com oAPI, você usa uma etapa para invocar `spark-submit` usando `command-runner.jar`.

Para obter mais informações sobre como enviar aplicações ao Spark, consulte o tópico [Submitting applications](#) na documentação do Apache Spark.

Para enviar uma etapa do Spark usando o console

1. Abra o EMR console da Amazon em <https://console.aws.amazon.com/emr>.
2. Em Cluster List (Lista de clusters), escolha o nome do cluster.
3. Role até a seção Steps (Etapas) e expanda-a. Em seguida, escolha Add step (Adicionar etapa).
4. Na caixa de diálogo Add Step (Adicionar etapa):
  - Para Step type (Tipo de etapa), escolha Spark application (Aplicativo Spark).
  - Para Name (Nome), aceite o nome padrão (aplicativo Spark) ou digite um novo nome.
  - Em Deploy mode (Modo de implantação), escolha o modo de Client (Cliente) ou de Cluster. O modo de cliente inicia o programa de driver na instância primária do cluster, enquanto o modo de cluster inicia o programa de driver no cluster. No modo cliente, a saída do registro do driver aparece nos registros das etapas, enquanto no modo cluster, a saída do registro do driver aparece nos registros do primeiro YARN contêiner. Para obter mais informações, consulte [Cluster mode overview](#) na documentação do Apache Spark.
  - Especifique as Spark-submit options (opções de Spark-submit) desejadas. Para obter mais informações sobre as opções de `spark-submit`, consulte [Launching applications with spark-submit](#).
  - Em Localização do aplicativo, especifique o URI caminho local ou S3 do aplicativo.
  - Para Arguments (Argumentos), deixe o campo em branco.
  - Para Action on failure (Ação na falha), aceite a opção padrão Continue (Continuar).
5. Escolha Adicionar. A etapa é exibida no console com o status Pendente.
6. O status da etapa muda de Pending (Pendente) para Running (Em execução) e depois para Completed (Concluído) conforme ela é executada. Para atualizar o status, escolha o ícone Refresh (Atualizar) acima da coluna Actions (Ações).
7. Os resultados da etapa estão localizados na página Detalhes do cluster do EMR console Amazon, ao lado da sua etapa, em Arquivos de log, se você tiver o registro configurado. Opcionalmente, você pode encontrar informações de etapas no bucket de logs configurado ao executar o cluster.

## Para enviar trabalhos para o Spark usando o AWS CLI

Envie uma etapa ao criar o cluster ou use o subcomando `aws emr add-steps` em um cluster existente.

1. Use `create-cluster`, conforme mostrado no exemplo a seguir.

### Note

Os caracteres de continuação de linha do Linux (`\`) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (`^`).

```
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
```

Como opção, você pode usar `command-runner.jar` conforme mostrado no exemplo a seguir.

```
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

Os caracteres de continuação de linha do Linux (`\`) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (`^`).

2. Como alternativa, adicione etapas a um cluster já em execução. Usar `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]
```

Como opção, você pode usar `command-runner.jar` conforme mostrado no exemplo a seguir.

```
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]
```

Para enviar trabalhos para o Spark usando o SDK for Java

1. O exemplo a seguir mostra como adicionar uma etapa a um cluster com o Spark usando 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. Visualize os resultados da etapa examinando os logs da etapa. Você pode fazer isso no AWS Management Console caso de ter ativado o registro, escolhendo Etapas, selecionando sua etapa e, em Arquivos de log, escolhendo `stdout` ou `stderr`. Para ver os logs disponíveis, escolha View Logs (Exibir logs).

## Substituir as definições de configuração padrão do Spark

Você pode querer substituir valores de configuração padrão do Spark para cada aplicativo. Você pode fazer isso ao enviar aplicativos usando uma etapa, que transmite essencialmente opções para `spark-submit`. Por exemplo, você pode querer alterar a memória alocada para um processo de executor modificando `spark.executor.memory`. Você poderia fornecer à opção `--executor-memory` um argumento semelhante ao seguinte:

```
spark-submit --executor-memory 1g --class org.apache.spark.examples.SparkPi /usr/lib/spark/examples/jars/spark-examples.jar 10
```

Da mesma forma, você pode ajustar `--executor-cores` e `--driver-memory`. Em uma etapa, você forneceria os seguintes argumentos para a etapa:

```
--executor-memory 1g --class org.apache.spark.examples.SparkPi /usr/lib/spark/examples/jars/spark-examples.jar 10
```

Você também pode ajustar as configurações que não possam não ter uma opção interna usando `--conf`. Para obter mais informações sobre outras configurações que são ajustáveis, consulte o tópico [Dynamically loading Spark properties](#) na documentação do Apache Spark.

## Visualizar o histórico de aplicações do Spark

Você pode visualizar os detalhes da interface do usuário do Spark, do YARN aplicativo e do Tez usando a guia Interfaces de usuário do aplicativo da página de detalhes de um cluster no console. As interfaces de usuário (UI) do EMR aplicativo Amazon facilitam a solução de problemas e a análise de trabalhos ativos e o histórico de trabalhos.

Para obter mais informações, consulte [Visualizar o histórico do aplicativo](#) no Amazon EMR Management Guide.

## Acesse a web do Spark UIs

Você pode visualizar a web do Spark UIs seguindo os procedimentos para criar um SSH túnel ou criar um proxy na seção [Connect to the cluster](#) no Amazon EMR Management Guide e, em seguida, navegando até o YARN ResourceManager para seu cluster. Escolha o link em Tracking UI (Interface do usuário de rastreamento) para o seu aplicativo. Se seu aplicativo estiver em execução, você verá ApplicationMaster. Isso o levará à interface do usuário web do aplicativo principal na porta 20888 onde o driver está localizado. O driver pode estar localizado no nó primário do cluster se você executar no modo YARN cliente. Se você estiver executando um aplicativo no modo de YARN cluster, o driver está localizado no do aplicativo no cluster. ApplicationMaster Se seu aplicativo tiver sido concluído, você verá o Histórico, que o levará ao número da porta da HistoryServer interface do usuário do Spark em 18080 do nó primário do EMR cluster. Isso é para aplicativos que já foram concluídos. Você também pode navegar até a HistoryServer interface do usuário do Spark diretamente em `http://master-public-dns-name:18080/`.

Com a EMR versão 5.25.0 da Amazon e versões posteriores, você pode acessar a interface do servidor de histórico do Spark a partir do console sem configurar um proxy web por meio de uma conexão. SSH Para obter mais informações, consulte [View Persistent Application User Interfaces](#).

## Usando o conector Amazon Kinesis Data Streams de streaming estruturado do Spark

As EMR versões 7.1.0 e posteriores da Amazon incluem um conector Amazon Kinesis Data Streams de streaming estruturado por Spark na imagem de lançamento. Com esse conector, você pode usar o Spark na Amazon EMR para processar dados armazenados no Amazon Kinesis Data Streams. O conector é compatível com os tipos de consumidor GetRecords (taxa de transferência compartilhada) e SubscribeToShard (fan-out aprimorado). Essa integração é baseada no [spark-sql-kinesis-connector](#). Para obter detalhes sobre como começar a usar o conector, consulte [README](#)o.

O exemplo a seguir demonstra como usar o conector para iniciar um aplicativo Spark com a Amazon. EMR

```
spark-submit my_kinesis_streaming_script.py
```

# Usando a integração do Amazon Redshift para o Apache Spark com a Amazon EMR

Com a EMR versão 6.4.0 e posterior da Amazon, cada imagem de lançamento inclui um conector entre o [Apache Spark](#) e o Amazon Redshift. Com esse conector, você pode usar o Spark na Amazon EMR para processar dados armazenados no Amazon Redshift. [Para as EMR versões 6.4.0 a 6.8.0 da Amazon, a integração é baseada no conector de código aberto. `spark-redshift`](#) Para as EMR versões 6.9.0 e posteriores da Amazon, a [integração do Amazon Redshift para o Apache Spark](#) foi migrada da versão comunitária para uma integração nativa.

## Tópicos

- [Inicialização de uma aplicação do Spark usando a integração do Amazon Redshift para Apache Spark](#)
- [Autenticação com a integração do Amazon Redshift para Apache Spark](#)
- [Leitura e gravação de e para o Amazon Redshift](#)
- [Considerações e limitações ao usar o conector do Spark](#)

## Inicialização de uma aplicação do Spark usando a integração do Amazon Redshift para Apache Spark

Para as EMR versões 6.4 a 6.9 da Amazon, você deve usar a `--packages` opção `--jars` ou para especificar quais dos seguintes JAR arquivos você deseja usar. A `--jars` opção especifica dependências armazenadas localmenteHDFS, em ou usando HTTP /S. Para ver outras localizações de arquivos suportadas pela `--jars` opção, consulte [Gerenciamento avançado de dependências](#) na documentação do Spark. A opção `--packages` especifica dependências armazenadas no repositório público Maven.

- `spark-redshift.jar`
- `spark-avro.jar`
- `RedshiftJDBC.jar`
- `minimal-json.jar`

As EMR versões 6.10.0 e superiores da Amazon não exigem a `minimal-json.jar` dependência e instalam automaticamente as outras dependências em cada cluster por padrão. Os exemplos a



seguintes mostram como iniciar uma aplicação do Spark com a integração do Amazon Redshift para Apache Spark.

#### Amazon EMR 6.10.0 +

O exemplo a seguir mostra como iniciar um aplicativo Spark com o `spark-redshift` conector com as EMR versões 6.10 e superiores da Amazon.

```
spark-submit my_script.py
```

#### Amazon EMR 6.4.0 - 6.9.x

Para iniciar um aplicativo Spark com o `spark-redshift` conector nas EMR versões 6.4 a 6.9 da Amazon, você deve usar a `--packages` opção `--jars` ou, como mostra o exemplo a seguir. Observe que os caminhos listados com a `--jars` opção são os caminhos padrão para os JAR arquivos.

```
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
```

## Autenticação com a integração do Amazon Redshift para Apache Spark

### Usando AWS Secrets Manager para recuperar credenciais e conectar-se ao Amazon Redshift

O exemplo de código a seguir mostra como você pode usar AWS Secrets Manager para recuperar credenciais para se conectar a um cluster do Amazon Redshift com a interface PySpark do Apache Spark em 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()
```

## Usando IAM para recuperar credenciais e conectar-se ao Amazon Redshift

Você pode usar o driver da JDBC versão 2 fornecido pelo Amazon Redshift para se conectar ao Amazon Redshift com o conector Spark. Para usar AWS Identity and Access Management (IAM), [configure sua IAM autenticação JDBC URL para usar](#). Para se conectar a um cluster do Redshift da AmazonEMR, você deve dar permissão à sua IAM função para recuperar credenciais temporárias. IAM atribua as seguintes permissões à sua IAM função para que ela possa recuperar credenciais e executar operações do Amazon S3.

- [Redshift: GetClusterCredentials](#) (para clusters provisionados do Amazon Redshift)
- [Redshift: DescribeClusters](#) (para clusters provisionados do Amazon Redshift)
- [Redshift: GetWorkgroup](#) (para grupos de trabalho sem servidor do Amazon Redshift)
- [Redshift: GetCredentials](#) (para grupos de trabalho sem servidor do Amazon Redshift)
- [s3: GetBucket](#)
- [s3: GetBucketLocation](#)
- [s3: GetObject](#)
- [s3: PutObject](#)
- [s3: GetBucketLifecycleConfiguration](#)

Para obter mais informações sobre `GetClusterCredentials`, consulte [Políticas de recursos para `GetClusterCredentials`](#).

Você também deve garantir que o Amazon Redshift possa assumir a IAM função durante COPY as UNLOAD operações.

```
{
 "Version": "2012-10-17",
 "Statement": [
 {
 "Effect": "Allow",
 "Principal": {
 "Service": "redshift.amazonaws.com"
 },
 "Action": "sts:AssumeRole"
 }
]
}
```

O exemplo a seguir usa IAM autenticação entre o Spark e o 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()
```

## Leitura e gravação de e para o Amazon Redshift

Os exemplos de código a seguir são usados PySpark para ler e gravar dados de amostra de e para um banco de dados do Amazon Redshift com fonte de dados API e usando o Spark. SQL

## Data source API

Use PySpark para ler e gravar dados de amostra de e para um banco de dados do Amazon Redshift com fonte de dados. 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

Use PySpark para ler e gravar dados de amostra de e para um banco de dados do Amazon Redshift com o 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()

```

## Considerações e limitações ao usar o conector do Spark

- Recomendamos que você ative a JDBC conexão do Spark na Amazon com o Amazon EMR Redshift. SSL
- Recomendamos que você gerencie as credenciais do cluster do Amazon Redshift no AWS Secrets Manager como uma prática recomendada. Consulte [Usando AWS Secrets Manager para recuperar credenciais para se conectar ao Amazon Redshift para](#) ver um exemplo.
- Recomendamos que você transmita uma IAM função com o parâmetro `aws_iam_role` para o parâmetro de autenticação do Amazon Redshift.
- Os `tempdir` URI pontos para uma localização do Amazon S3. Esse diretório temporário não é limpo automaticamente e, portanto, pode incorrer em custos adicionais.
- Considere as seguintes recomendações para o Amazon Redshift:
  - Recomendamos bloquear o acesso público ao cluster do Amazon Redshift.

- Recomendamos ativar o [registro em log de auditoria do Amazon Redshift](#).
- Recomendamos que você ative a [Criptografia em repouso do Amazon Redshift](#).
- Considere as seguintes recomendações para o Amazon S3:
  - Recomendamos que você [bloqueeie o acesso público aos buckets do Amazon S3](#).
  - Recomendamos que você use [criptografia no lado do servidor do Amazon S3](#) para criptografar os buckets do Amazon S3 usados.
  - Recomendamos que você use as [políticas de ciclo de vida do Amazon S3](#) para definir as regras de retenção para o bucket do Amazon S3.
- A Amazon EMR sempre verifica o código importado do código aberto para a imagem. Por motivos de segurança, não oferecemos suporte aos seguintes métodos de autenticação do Spark para o Amazon S3:
  - Definindo chaves de AWS acesso na classificação `hadoop-env` de configuração
  - Teclas de AWS acesso de codificação no `tempdir` URI

Para obter mais informações sobre como usar o conector e os parâmetros compatíveis, consulte os seguintes recursos:

- [Integração do Amazon Redshift para Apache Spark](#) no Guia de gerenciamento do Amazon Redshift.
- O [repositório da comunidade spark-redshift](#) no GitHub.

## Histórico de versões do Spark

A tabela a seguir lista a versão do Spark incluída em cada versão de lançamento da AmazonEMR, junto com os componentes instalados com o aplicativo. Para obter as versões dos componentes em cada versão, consulte a seção Versão de componentes da sua versão em [Versões de lançamento do Amazon EMR 7.x](#), [Versões de lançamento do Amazon EMR 6.x](#) ou [Versões de lançamento do Amazon EMR 5.x](#).

### Important

[A versão 2.3.1 do Apache Spark, disponível a partir da EMR versão 5.16.0 da Amazon, endereça -2018-8024 e -2018-1334. CVE CVE](#) Recomendamos que você migre as versões anteriores do Spark para a versão 2.3.1 ou posteriores.

## Informações das versões do Spark

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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, 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-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-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-7.1.0                    | 3.5.0           | delta, emrfs, emr-goodies, emr-ddb, emr-s3-select,                                                                                                                                                                                                                                                                                                                                           |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o Spark                                                                                                                                                                                                                                                                                                                                                  |
|------------------------------|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                              |                 | 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 |



| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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-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-6.14.0                   | 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-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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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-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-6.12.0                   | 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-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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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-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-6.10.0                   | 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-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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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-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-6.9.0                    | 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-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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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-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-6.8.0                    | 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-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, iceberg, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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-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-5.36.1                   | 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-resourcemanager, hadoop-yarn-timeline-server, hudi, hudi-spark, livy-server, nginx, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave          |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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 |



| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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-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-6.5.0                    | 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-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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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-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-6.3.1                    | 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-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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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             |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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-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             |
| 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-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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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-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-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-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 |



| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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 |
| emr-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                   |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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 |



| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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-https-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-https-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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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-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-5.16.1                   | 2.3.1           | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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                |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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 |



| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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-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-5.14.2                   | 2.3.0           | aws-sagemaker-spark-sdk, emrfs, emr-goodies, emr-ddb, hadoop-client, hadoop-hdfs-datanode, hadoop-hdfs-library, hadoop-hdfs-namenode, hadoop-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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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                          |



| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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              |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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-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-5.8.3                    | 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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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-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-5.7.1                    | 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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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                                       |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o Spark                                                                                                                                                                                                                                    |
|------------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.5.2                    | 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-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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.5.0                    | 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, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |



| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o Spark                                                                                                                                                                                                                                    |
|------------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.3.1                    | 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, 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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |
| emr-5.2.3                    | 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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o Spark                                                                                                                                                                                                                                    |
|------------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-4.9.6                    | 1.6.3           | 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, 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-https-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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o Spark                                                                                                                                                                                                                                    |
|------------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-4.8.5                    | 1.6.3           | 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, 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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, 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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o Spark                                                                                                                                                                                                                                    |
|------------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-4.8.2                    | 1.6.2           | 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-resourcemanager, 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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, 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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave |



| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o Spark                                                                                                                                                                                                                                     |
|------------------------------|-----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-4.7.4                    | 1.6.2           | 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 |
| emr-4.7.3                    | 1.6.2           | 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 |
| emr-4.7.2                    | 1.6.2           | 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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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                    |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o Spark                                                                                                                                                                                                                 |
|------------------------------|-----------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-4.3.0                    | 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 |
| emr-4.2.0                    | 1.5.2           | 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-4.1.0                    | 1.5.0           | 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 |

| Gravadora Amazon EMR Release | Versão do Spark | Componentes instalados com o 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

O Apache Sqoop é uma ferramenta para transferir dados entre Amazon S3, Hadoop e bancos de dados. HDFS RDBMS Para obter mais informações, consulte o [site do Apache Sqoop](#). O Sqoop está incluído nas EMR versões 5.0.0 e posteriores da Amazon. As versões anteriores incluem o Sqoop como uma aplicação em sandbox. Para ter mais informações, consulte [Versões de lançamento do Amazon EMR 4.x](#).

## Tópicos

- [Informações de versão do Sqoop](#)
- [Considerações sobre o Sqoop na Amazon EMR](#)
- [Histórico de versões do Sqoop](#)

## Informações de versão do Sqoop

### Versão Sqoop para 7.2.0

A tabela a seguir lista a versão do Sqoop incluída na versão mais recente da série Amazon EMR 7.x, junto com os componentes que a Amazon EMR instala com o Sqoop.

Para a versão dos componentes instalados com o Sqoop nesta versão, consulte Versões de componentes da [versão 7.2.0](#).

### Informações sobre a versão do Sqoop para o emr-7.2.0

| Gravadora EMR de lançamento da Amazon | Versão do Sqoop | Componentes instalados com o 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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn- |

| Gravadora EMR de lançamento da Amazon | Versão do Sqoop | Componentes instalados com o Sqoop            |
|---------------------------------------|-----------------|-----------------------------------------------|
|                                       |                 | timeline-server, mariadb-server, sqoop-client |

### Versão do Sqoop para 6.15.0

A tabela a seguir lista a versão do Sqoop incluída na versão mais recente da série Amazon EMR 6.x, junto com os componentes que a Amazon EMR instala com o Sqoop.

Para obter a versão dos componentes instalados com o Sqoop nessa versão, consulte [Release 6.15.0 Component Versions](#).

### Informações de versão do Sqoop para o emr-6.15.0

| Gravadora EMR de lançamento da Amazon | Versão do Sqoop | Componentes instalados com o 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 |

### Versão Sqoop para 5.36.2

A tabela a seguir lista a versão do Sqoop incluída na versão mais recente da série Amazon EMR 5.x, junto com os componentes que a Amazon EMR instala com o Sqoop.

Para a versão dos componentes instalados com o Sqoop nesta versão, consulte [Versões de componentes da versão 5.36.2](#).

## Informações sobre a versão do Sqoop para o emr-5.36.2

| Gravadora EMR de lançamento da Amazon | Versão do Sqoop | Componentes instalados com o 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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |

## Considerações sobre o Sqoop na Amazon EMR

Considere os seguintes itens ao executar o Sqoop na Amazon EMR.

### Usando o Sqoop com integração HCatalog

O Sqoop na Amazon EMR oferece suporte à integração com o [Sqoop](#). HCatalog Ao usar o Sqoop para gravar a saída em uma HCatalog tabela no Amazon S3, desative a gravação direta da EMR Amazon definindo `mapred.output.direct.NativeS3FileSystem` as `mapred.output.direct.EmrFileSystem` propriedades e como `false`. Para ter mais informações, consulte [Usando HCatalog](#). Você pode usar os comandos `-D mapred.output.direct.NativeS3FileSystem=false` e `-D mapred.output.direct.EmrFileSystem=false` do Hadoop. Se você não desabilitar a gravação direta, nenhum erro ocorrerá, mas a tabela será criada no Amazon S3 e nenhum dado será gravado.

### Suporte para Sqoop JDBC e banco de dados

Por padrão, o Sqoop tem um driver MariaDB e Postgre instalado. SQL O SQL driver Postgre instalado para o Sqoop funciona somente para o Postgre 8.4. SQL Para instalar um conjunto



alternativo de JDBC conectores para o Sqoop, conecte-se ao nó principal do cluster e instale-os nele. `/usr/lib/sqoop/lib` A seguir estão os links para vários JDBC conectores:

- MariaDB: [Sobre o MariaDB Connector/J](#).
- PostgreSQL: driver do [Postger SQL JDBC](#).
- SQLServer: [Baixe o JDBC driver da Microsoft para o SQL servidor](#).
- MeuSQL: [Conector de download/J](#)
- Oracle: [obtenha JDBC drivers Oracle e UCP do repositório Oracle Maven](#)

Os bancos de dados compatíveis com o Sqoop estão listados no seguinte url, [http://sqoop.apache.org/docs/VERSION/SqoopUserGuide.html#\\_supported\\_databases](http://sqoop.apache.org/docs/VERSION/SqoopUserGuide.html#_supported_databases), onde *version* é a versão do Sqoop que você está usando, por exemplo, 1.4.6. Se a cadeia de JDBC conexão não corresponder à dessa lista, você deverá especificar um driver.

Por exemplo, você pode exportar para uma tabela de banco de dados do Amazon Redshift com o seguinte comando (para JDBC 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
```

Você pode usar as cadeias de conexão MariaDB e SQL My, mas se especificar a string de conexão MariaDB, precisará especificar o driver:

```
sqoop export --connect jdbc:mariadb://$HOSTNAME:3306/mydb --table mysqoopexport
--export-dir s3://mybucket/myinputfiles/ --driver org.mariadb.jdbc.Driver --
username master --password Mymasterpass1
```

Se você estiver usando a criptografia Secure Socket Layer para acessar seu banco de dados, precisará usar JDBC URI algo semelhante no seguinte exemplo de exportação do Sqoop:

```
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
```

Para obter mais informações sobre SSL criptografia emRDS, consulte Como [usar SSL para criptografar uma conexão com uma instância de](#) banco de dados no Guia do RDS usuário da Amazon.

Para obter mais informações, consulte a documentação do [Apache Sqoop](#).

## Proteção da senha

Há vários métodos que você pode escolher para transmitir sua senha com segurança:

### Java KeyStore

O método preferido criptografa a senha com um Java KeyStore (JKS), eliminando a necessidade de armazenar a senha em um formato legível.

1. Crie um alias de senha. Quando solicitado, insira a senha usada para acessar o banco de dados.

```
hadoop credential create mydb.password.alias -provider jceks://hdfs/user/root/
mysql.password.jceks
```

2. Use o alias de senha para executar o trabalho do 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
```

### --password-file

É possível usar o comando `--password-file` para transmitir a senha por meio de um arquivo, conforme mostrado no seguinte exemplo:

1. Crie um arquivo que contenha a senha:

```
echo -n 'Mymasterpass1' > /home/hadoop/mysql-pass.password
```

2. Use o arquivo para executar o trabalho do 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

É possível usar o comando -P para inserir a senha por meio de um prompt, conforme mostrado no seguinte exemplo:

```
sqoop export --connect jdbc:mariadb://$HOSTNAME:3306/mydb --table mysqoopexport --
export-dir s3://mybucket/myinputfiles/ --driver org.mariadb.jdbc.Driver --username
master -P
```

## Histórico de versões do Sqoop

A tabela a seguir lista a versão do Sqoop incluída em cada versão de lançamento da AmazonEMR, junto com os componentes instalados com o aplicativo. Para obter as versões dos componentes em cada versão, consulte a seção Versão de componentes da sua versão em [Versões de lançamento do Amazon EMR 7.x](#), [Versões de lançamento do Amazon EMR 6.x](#) ou [Versões de lançamento do Amazon EMR 5.x](#).

### Informações de versão do Sqoop

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o 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- |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                  |
|------------------------------|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                              |                 | 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-hdfs-namenode, hadoop-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, mariadb-server, sqoop-client |
| emr-7.1.0                    | 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                  |
|------------------------------|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-7.0.0                    | 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-6.15.0                   | 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                  |
|------------------------------|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                  |
|------------------------------|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                  |
|------------------------------|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |



| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                  |
|------------------------------|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                  |
|------------------------------|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                  |
|------------------------------|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                  |
|------------------------------|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                  |
|------------------------------|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                  |
|------------------------------|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                  |
|------------------------------|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                  |
|------------------------------|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |



| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                  |
|------------------------------|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                  |
|------------------------------|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                  |
|------------------------------|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                  |
|------------------------------|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                  |
|------------------------------|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                  |
|------------------------------|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                |
|------------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                |
|------------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |



| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                |
|------------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                |
|------------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                |
|------------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                |
|------------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                |
|------------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                |
|------------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                |
|------------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                |
|------------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |



| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                |
|------------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                |
|------------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                |
|------------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                |
|------------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                |
|------------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                |
|------------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                |
|------------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                |
|------------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |



| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                |
|------------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                |
|------------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                |
|------------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                |
|------------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                |
|------------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                |
|------------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                |
|------------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                |
|------------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |



| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                                                |
|------------------------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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                              |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                    |
|------------------------------|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                    |
|------------------------------|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                    |
|------------------------------|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                    |
|------------------------------|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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-httpfs-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-httpfs-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                   |
|------------------------------|-----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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-https-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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, mysql-server, sqoop-client |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                    |
|------------------------------|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Sqoop | Componentes instalados com o Sqoop                                                                                                                                                                                                                    |
|------------------------------|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resourcesanager, mysql-server, sqoop-client |



# TensorFlow

TensorFlow é uma biblioteca matemática simbólica de código aberto para aplicativos de inteligência de máquina e aprendizado profundo. Para obter mais informações, consulte o [TensorFlow site](#). TensorFlow está disponível com a EMR versão 5.17.0 e posterior da Amazon.

A tabela a seguir lista a versão TensorFlow incluída na versão mais recente da série Amazon EMR 7.x, junto com os componentes com os quais a Amazon é EMR instalada. TensorFlow

Para a versão dos componentes instalados TensorFlow nesta versão, consulte [Versões de componentes da versão 7.2.0](#).

TensorFlow informações sobre a versão do emr-7.2.0

| Gravadora EMR de lançamento da Amazon | TensorFlow Versão | Componentes instalados com 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 |

A tabela a seguir lista a versão TensorFlow incluída na versão mais recente da série EMR 6.x da Amazon, junto com os componentes com os quais a Amazon é EMR instalada. TensorFlow

Para a versão dos componentes instalados TensorFlow nesta versão, consulte [Versões de componentes da versão 6.15.0](#).

## TensorFlow informações da versão do emr-6.15.0

| Gravadora EMR de lançamento da Amazon | TensorFlow Versão | Componentes instalados com 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 |

A tabela a seguir lista a versão TensorFlow incluída na versão mais recente da série Amazon EMR 5.x, junto com os componentes com os quais a Amazon é EMR instalada. TensorFlow

Para a versão dos componentes instalados TensorFlow nesta versão, consulte Versões de componentes da [versão 5.36.2](#).

## TensorFlow informações da versão do emr-5.36.2

| Gravadora EMR de lançamento da Amazon | TensorFlow Versão | Componentes instalados com 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 compilações por tipo de EC2 instância da Amazon

A Amazon EMR usa diferentes versões da TensorFlow biblioteca, dependendo dos tipos de instância que você escolher para o seu cluster. A tabela a seguir lista as compilações por tipo de instância.

| EC2tipos de instância | TensorFlow construir                                                                                                                                                                                                                                                                                              |
|-----------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| M5 e C5               | Tensorflow 1.9.0 com otimização Intel MKL                                                                                                                                                                                                                                                                         |
| P2                    | Tensorflow 1.9.0 com 9.2, com CUDA 7.1 DNN                                                                                                                                                                                                                                                                        |
| P3                    | Tensorflow 1.9.0 com CUDA 9.2, cu 7.1, 2.2.13 DNN NCCL<br><br><a href="#">A Nvidia NCCL</a> está disponível somente em instâncias P3. Contrato de licença de usuário final (EULA): <a href="#">Ao usar componentes da Nvidia na AmazonEMR, você concorda com os termos e condições descritos no produto. EULA</a> |
| Todos os outros       | Tensorflow 1.9.0                                                                                                                                                                                                                                                                                                  |

## Segurança

Além de seguir as orientações em [Como usar TensorFlow com segurança](#), recomendamos que você execute seu cluster em uma sub-rede privada para ajudá-lo a limitar o acesso a fontes confiáveis. Para obter mais informações, consulte [VPCas opções da Amazon](#) no Guia EMR de gerenciamento da Amazon.

## Usando TensorBoard

TensorBoard é um conjunto de ferramentas de visualização para TensorFlow programas. Para obter mais informações, consulte [TensorBoard: Aprendizado visualizado](#) no site do Tensorflow.

Para usar TensorBoard com a AmazonEMR, você deve começar TensorBoard no nó principal do cluster.

## Para usar o tensorboard com o Tensorflow na Amazon EMR

1. Conecte-se ao nó principal do cluster usando SSH. Para obter mais informações, consulte [Conecte-se ao nó principal usando SSH](#) o Amazon EMR Management Guide.
2. Digite o seguinte comando para iniciar o Tensorboard no nó principal. Substitua `/my/log/directory` por um diretório no nó principal o qual você tiver gerado e armazenado os dados de resumo usando um gravador de resumo.

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
```

Por padrão, o nó principal hospeda TensorBoard usando a porta 6006 e o DNS nome público principal. Depois de iniciar TensorBoard, a saída da linha de comando apresenta o URL que pode ser usado para se conectar TensorBoard, conforme mostrado no exemplo a seguir:

```
TensorBoard 1.9.0 at http://master-public-dns-name:6006 (Press CTRL+C to quit)
```

3. Configure o acesso a interfaces web no nó principal a partir de clientes confiáveis. Para obter mais informações, consulte [Exibir interfaces web hospedadas em EMR clusters](#) da Amazon no Amazon EMR Management Guide.
4. Aberto TensorBoard em `http://master-public-dns-name:6006`.

## TensorFlow histórico de lançamentos

A tabela a seguir lista a versão TensorFlow incluída em cada versão de lançamento da AmazonEMR, junto com os componentes instalados com o aplicativo. Para obter as versões dos componentes em cada versão, consulte a seção Versão de componentes da sua versão em [Versões de lançamento do Amazon EMR 7.x](#), [Versões de lançamento do Amazon EMR 6.x](#) ou [Versões de lançamento do Amazon EMR 5.x](#).

## TensorFlow informações sobre a versão

| Gravadora Amazon EMR Release | TensorFlow Versão | Componentes instalados com 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-                                                                                                                                                     |

| Gravadora Amazon EMR Release | TensorFlow Versão | Componentes instalados com TensorFlow                                                                                                                                                                                                                                       |
|------------------------------|-------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                              |                   | hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, tensorflow                                                                                                 |
| emr-6.15.0                   | 2.11.0            | emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, tensorflow |
| emr-6.14.0                   | 2.11.0            | emrfs, emr-goodies, hadoop-<br>client, hadoop-hdfs-datanode,<br>hadoop-hdfs-library, hadoop-<br>hdfs-namenode, hadoop-<br>httpfs-server, hadoop-kms-<br>server, hadoop-yarn-nodema-<br>nager, hadoop-yarn-resour-<br>cemanager, hadoop-yarn-<br>timeline-server, tensorflow |

| Gravadora Amazon EMR Release | TensorFlow Versão | Componentes instalados com 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 |

| Gravadora Amazon EMR Release | TensorFlow Versão | Componentes instalados com TensorFlow                                                                                                                                                                                                     |
|------------------------------|-------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.11.0                   | 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-6.10.1                   | 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-6.10.0                   | 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 |



| Gravadora Amazon EMR Release | TensorFlow Versão | Componentes instalados com 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 |

| Gravadora Amazon EMR Release | TensorFlow Versão | Componentes instalados com TensorFlow                                                                                                                                                                                                     |
|------------------------------|-------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.8.0                    | 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-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-https-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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, tensorflow |

| Gravadora Amazon EMR Release | TensorFlow Versão | Componentes instalados com TensorFlow                                                                                                                                                                                                      |
|------------------------------|-------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.36.0                   | 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-6.6.0                    | 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-5.35.0                   | 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 |

| Gravadora Amazon EMR Release | TensorFlow Versão | Componentes instalados com TensorFlow                                                                                                                                                                                                     |
|------------------------------|-------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.5.0                    | 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-6.4.0                    | 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-6.3.1                    | 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 |

| Gravadora Amazon EMR Release | TensorFlow Versão | Componentes instalados com 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-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, tensorflow |

| Gravadora Amazon EMR Release | TensorFlow Versão | Componentes instalados com 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 |

| Gravadora Amazon EMR Release | TensorFlow Versão | Componentes instalados com 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 |

| Gravadora Amazon EMR Release | TensorFlow Versão | Componentes instalados com 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, tensorflow |



| Gravadora Amazon EMR Release | TensorFlow Versão | Componentes instalados com 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-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, tensorflow |

| Gravadora Amazon EMR Release | TensorFlow Versão | Componentes instalados com 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 |
| emr-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 |

| Gravadora Amazon EMR Release | TensorFlow Versão | Componentes instalados com 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 |

| Gravadora Amazon EMR Release | TensorFlow Versão | Componentes instalados com 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 |

| Gravadora Amazon EMR Release | TensorFlow Versão | Componentes instalados com 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 |

| Gravadora Amazon EMR Release | TensorFlow Versão | Componentes instalados com TensorFlow                                                                                                                                                                                                     |
|------------------------------|-------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.23.0                   | 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-5.22.0                   | 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-5.21.2                   | 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 |

| Gravadora Amazon EMR Release | TensorFlow Versão | Componentes instalados com 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, tensorflow |

| Gravadora Amazon EMR Release | TensorFlow Versão | Componentes instalados com TensorFlow                                                                                                                                                                                                     |
|------------------------------|-------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.20.0                   | 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-5.19.1                   | 1.11.0            | 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-5.19.0                   | 1.11.0            | 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 |



| Gravadora Amazon EMR Release | TensorFlow Versão | Componentes instalados com 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-resource-manager, 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-resource-manager, 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-resource-manager, hadoop-yarn-timeline-server, tensorflow |

| Gravadora Amazon EMR Release | TensorFlow Versão | Componentes instalados com 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

O Apache Tez é uma estrutura que cria um gráfico acíclico direcionado complexo (DAG) de tarefas para processamento de dados. Você pode usá-lo como uma alternativa ao Hadoop MapReduce para alguns casos de uso. Por exemplo, você pode executar fluxos de trabalho do Pig e do Hive com o Hadoop MapReduce ou usar o Tez como um mecanismo de execução. Para obter mais informações, consulte <https://tez.apache.org/>. As EMR versões 4.7.0 e superiores da Amazon incluem o Tez.

A tabela a seguir lista a versão do Tez incluída na versão mais recente da série Amazon EMR 7.x, junto com os componentes que a Amazon EMR instala com o Tez.

Para a versão dos componentes instalados com o Tez nesta versão, consulte [Versões de componentes da versão 7.2.0](#).

Informações sobre a versão Tez para emr-7.2.0

| Gravadora EMR de lançamento da Amazon | Versão do Tez | Componentes instalados com o 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 |

A tabela a seguir lista a versão do Tez incluída na versão mais recente da série EMR 6.x da Amazon, junto com os componentes que a Amazon EMR instala com o Tez.

Para obter a versão dos componentes instalados com o Tez nessa versão, consulte [Release 6.15.0 Component Versions](#).

## Informações de versão do Tez para o emr-6.15.0

| Gravadora EMR de lançamento da Amazon | Versão do Tez | Componentes instalados com o 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 |

A tabela a seguir lista a versão do Tez incluída na versão mais recente da série Amazon EMR 5.x, junto com os componentes que a Amazon EMR instala com o Tez.

Para a versão dos componentes instalados com o Tez nesta versão, consulte [Versões de componentes da versão 5.36.2](#).

## Informações sobre a versão Tez do emr-5.36.2

| Gravadora EMR de lançamento da Amazon | Versão do Tez | Componentes instalados com o 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 |

## Tópicos

- [Criar um cluster com o Tez](#)
- [Configurar o Tez](#)
- [Interface do usuário da Web do Tez](#)
- [Timeline Server](#)
- [Histórico de versões do Tez](#)

## Criar um cluster com o Tez

Para instalar o Tez, escolha o Apache Tez como uma aplicação quando criar o cluster.

Para criar um cluster com o Tez instalado usando o console

1. Navegue até o novo EMR console da Amazon e selecione Alternar para o console antigo na navegação lateral. Para obter mais informações sobre o que esperar ao alternar para o console antigo, consulte [Usar o console antigo](#).
2. Escolha Create cluster (Criar cluster), Go to advanced options (Ir para opções avançadas).
3. Em Configuração de software, selecione uma Versão do emr-4.7.0 ou versões posteriores.
4. Selecione Tez junto com outros aplicativos que você deseja que EMR a Amazon instale.
5. Selecione outras opções conforme necessário e escolha Create cluster (Criar cluster).

Para criar um cluster com o Tez usando o AWS CLI

- Use o comando `create-cluster` com a opção `--applications` para especificar o Tez. O exemplo a seguir cria um cluster com o Tez instalado.

### Note

Os caracteres de continuação de linha do Linux (`\`) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (`^`).

```
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
```

## Configurar o Tez

Você pode personalizar o Tez definindo valores com a classificação de configuração `tez-site`, que define as configurações no arquivo de configuração `tez-site.xml`. Para obter mais informações, consulte a [TezConfiguration](#) documentação do Apache Tez. Para alterar o Hive ou o Pig para usar o mecanismo de execução do Tez, use o `hive-site` e classificações de configuração de `pig-properties`, conforme apropriado. Os exemplos são mostrados abaixo.

## Exemplo de configuração

Example Exemplo: personalizar o nível de registro raiz do Tez e configurar o Tez como mecanismo de execução para o Hive e o Pig

O comando de exemplo `create-cluster` mostrado a seguir cria um cluster com o Tez, o Hive e o Pig instalados. O comando faz referência a um arquivo armazenado no Amazon S3, `myConfig.json`, que especifica as propriedades da classificação `tez-site` que define `tez.am.log.level` como `DEBUG` e define o mecanismo de execução do Tez para Hive e Pig usando as classificações de configuração `hive-site` e `pig-properties`.

### Note

Os caracteres de continuação de linha do Linux (`\`) são incluídos para facilitar a leitura. Eles podem ser removidos ou usados em comandos do Linux. No Windows, remova-os ou substitua-os por um sinal de interpolação (`^`).

```
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
```

Os conteúdos de exemplo de `myConfig.json` são mostrados abaixo.

```
[
 {
```

```

 "Classification": "tez-site",
 "Properties": {
 "tez.am.log.level": "DEBUG"
 }
 },
 {
 "Classification": "hive-site",
 "Properties": {
 "hive.execution.engine": "tez"
 }
 },
 {
 "Classification": "pig-properties",
 "Properties": {
 "exectype": "tez"
 }
 }
]

```

### Note

Com a Amazon EMR versão 5.21.0 e posterior, você pode substituir as configurações do cluster e especificar classificações de configuração adicionais para cada grupo de instâncias em um cluster em execução. Você faz isso usando o EMR console da Amazon, o AWS Command Line Interface (AWS CLI) ou AWS SDK o. Para obter mais informações, consulte [Supplying a Configuration for an Instance Group in a Running Cluster](#).

## Abertura de divisão assíncrona do Tez

Quando há um grande número de arquivos pequenos no caminho da tabela e uma consulta tenta ler todos eles, cada arquivo pequeno que corresponde a cada divisão individual é combinado em uma divisão agrupada do Tez. Então, um único mapper processa a única divisão agrupada do Tez. Como a execução é síncrona, cada divisão individual sob a divisão agrupada é processada uma a uma. Isso exige que os objetos de `RecordReader` processem as divisões de forma síncrona.

| Nome                                                          | Classificação | Descrição                                      |
|---------------------------------------------------------------|---------------|------------------------------------------------|
| <code>tez.grouping.split</code><br><code>.init.threads</code> | tez-site      | Especifica o número de threads do daemon que o |

| Nome                                               | Classificação         | Descrição                                                                                                                                                                                                                                                                                                                                                                                               |
|----------------------------------------------------|-----------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                    |                       | Tez usa para pré-iniciar o <code>RecordReaders</code> e abrir as divisões. Em tabelas ACID, o valor máximo compatível de <code>tez.grouping.split.init.threads</code> é 1.                                                                                                                                                                                                                              |
| <code>tez.grouping.split.init.recordreaders</code> | <code>tez-site</code> | Especifica o número de <code>RecordReaders</code> a serem mantidos pré-inicializados por threads do daemon. Isso pode ajudar quando a divisão agrupada do Tez contém um grande número de <code>InputSplits</code> . A inicialização de <code>RecordReaders</code> para processar essas divisões de entrada pode ser feita de forma assíncrona com threads do daemon em vez de processamento sequencial. |

## Avaliação comparativa de abertura de divisão assíncrona do Tez

Usamos os ambientes e configurações a seguir para comparar a capacidade de abertura de divisão assíncrona do Tez.

- Ambiente de referência — EMR cluster da Amazon com 1 nó primário que usa `m5.16xlarge` e 16 nós principais que usam `m5.16xlarge`.
- Configurações de avaliação comparativa: para simular o cenário de avaliação comparativa em que um grande número de divisões de entrada estão em uma única divisão agrupada do Tez, `tez.grouping.split-count` é definido como 1.
- Tabela usada para avaliação comparativa: a tabela contém 200 partições, com cada partição contendo um único arquivo. A avaliação comparativa é feita para quando essa tabela contiver



arquivos CSV e quando essa tabela contiver arquivos Parquet. Consulta do Hive para avaliação comparativa: `SELECT COUNT(*)` da tabela dez vezes e calcule o runtime médio.

- Configurações para habilitar a abertura de divisão assíncrona do Tez: da seguinte forma:
  - `tez.grouping.split.init.threads = 4`
  - `tez.grouping.split.init.recordreaders = 10`

| Conjunto de dados         | Recurso desabilitado (linha de base) | Recurso habilitado | Melhoria |
|---------------------------|--------------------------------------|--------------------|----------|
| Conjunto de dados CSV     | 90.26 segundos                       | 79.20 segundos     | 12,25%   |
| Conjunto de dados Parquet | 54.67 segundos                       | 42.23 segundos     | 22,75%   |

## Interface do usuário da Web do Tez

O Tez tem sua própria interface de usuário Web. Para visualizar a interface do usuário da web, consulte o seguinte URL.

```
http://masterDNS:8080/tez-ui
```

Para habilitar a guia Consultas do Hive na interface do usuário da Web do Tez, defina a configuração a seguir.

```
[
 {
 "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"
 }
 }
]
```

Você também pode visualizar os detalhes do Tez, do Spark e da interface do usuário do YARN aplicativo usando links na guia Interfaces de usuário do aplicativo da página de detalhes de um cluster no console. As interfaces de usuário (UI) de EMR aplicativos da Amazon são hospedadas fora do cluster e estão disponíveis após o encerramento do cluster. Eles não exigem que você configure uma SSH conexão ou proxy da web, facilitando a solução de problemas e a análise de trabalhos ativos e histórico de trabalhos.

Para obter mais informações, consulte [Visualizar o histórico do aplicativo](#) no Amazon EMR Management Guide.

## Timeline Server

O YARN Timeline Server está configurado para ser executado quando o Tez é instalado. Para visualizar trabalhos enviados por meio do Tez ou de mecanismos de MapReduce execução usando o Timeline Server, visualize a interface do usuário da web usando o URL `http://master-public-DNS:8188` Para obter mais informações, consulte [Visualizar interfaces web hospedadas em EMR clusters da Amazon](#) no Amazon EMR Management Guide.

## Histórico de versões do Tez

A tabela a seguir lista a versão do Tez incluída em cada versão de lançamento da AmazonEMR, junto com os componentes instalados com o aplicativo. Para obter as versões dos componentes em cada versão, consulte a seção Versão de componentes da sua versão em [Versões de lançamento do Amazon EMR 7.x](#), [Versões de lançamento do Amazon EMR 6.x](#) ou [Versões de lançamento do Amazon EMR 5.x](#).

### Informações de versão do Tez

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o Tez                                                                                                                                                  |
|------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-7.2.0                    | 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-resour |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o Tez                                                                                                                                                                                                                     |
|------------------------------|---------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                              |               | cemanager, 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-resource-manager, hadoop-yarn-timeline-server, tez-on-yarn                |
| emr-7.1.0                    | 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-resource-manager, hadoop-yarn-timeline-server, tez-on-yarn, tez-on-worker |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o Tez                                                                                                                                                                                                                    |
|------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-7.0.0                    | 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 |
| emr-6.15.0                   | 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 |
| emr-6.14.0                   | 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o Tez                                                                                                                                                                                                                    |
|------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.13.0                   | 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 |
| emr-6.12.0                   | 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 |
| emr-6.11.1                   | 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o Tez                                                                                                                                                                                                                    |
|------------------------------|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.11.0                   | 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 |
| emr-6.10.1                   | 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 |
| emr-6.10.0                   | 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o Tez                                                                                                                                                                                                     |
|------------------------------|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.9.1                    | 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 |
| emr-6.9.0                    | 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |



| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |
| emr-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 |



| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |



| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |



| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |



| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |

| Gravadora Amazon EMR Release | Versão do Tez | Componentes instalados com o 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 |

## Notas da versão do Tez por versão

### Tópicos

- [Amazon EMR 7.1.0 - Notas de lançamento do Tez](#)
- [Amazon EMR 6.15.0 - Notas de lançamento do Tez](#)
- [Amazon EMR 6.14.0 - Notas de lançamento do Tez](#)
- [Amazon EMR 6.13.0 - Notas de lançamento do Tez](#)
- [Amazon EMR 6.12.0 - Notas de lançamento do Tez](#)
- [Amazon EMR 6.11.0 - Notas de lançamento do Tez](#)
- [Amazon EMR 6.10.0 - Notas de lançamento do Tez](#)
- [Amazon EMR 6.9.0 - Notas de lançamento do Tez](#)
- [Amazon EMR 6.8.0 - Notas de lançamento do Tez](#)
- [Amazon EMR 6.7.0 - Notas de lançamento do Tez](#)
- [Amazon EMR 6.6.0 - Notas de lançamento do Tez](#)

## Amazon EMR 7.1.0 - Notas de lançamento do Tez

### Amazon EMR 7.1.0 - Mudanças de chá

| Tipo             | Descrição                                                                             |
|------------------|---------------------------------------------------------------------------------------|
| Correção de bugs | <a href="#">TEZ-4394</a> — Netty4 ShuffleHandler: deve usar 1 thread principal (#256) |
| Upgrade          | Atualiza o Netty para 4.1.100.Final.                                                  |
| Upgrade          | Atualiza o Jetty para 9.4.53.v20231009.                                               |

## Amazon EMR 6.15.0 - Notas de lançamento do Tez

### Amazon EMR 6.15.0 - Mudanças no chá

| Tipo     | Descrição                                                                 |
|----------|---------------------------------------------------------------------------|
| Atributo | <a href="#">TEZ-4397</a> : O Open Tez Input se divide de forma assíncrona |

| Tipo    | Descrição                                                      |
|---------|----------------------------------------------------------------|
| Upgrade | <a href="#">TEZ-4493</a> : Atualize o Apache Hadoop para 3.3.6 |

## Amazon EMR 6.15.0 - Características do Tez

- [Abertura de divisão assíncrona Tez](#) — O Amazon EMR 6.15.0 introduz configurações que você pode especificar para abrir de forma assíncrona as divisões de entrada em uma divisão agrupada Tez. O recurso foi iniciado por [TEZ-4397](#), mas teve regressões no Hive. O Amazon EMR Hive corrigiu as regressões e outros bugs na tabela do HiveACID. Essa melhoria resulta em uma performance mais rápida das consultas de leitura quando há um grande número de divisões de entrada em uma única divisão agrupada do Tez. Para ter mais informações, consulte [Abertura de divisão assíncrona do Tez](#).

## Amazon EMR 6.14.0 - Notas de lançamento do Tez

### Amazon EMR 6.14.0 - Mudanças no chá

| Tipo     | Descrição                             |
|----------|---------------------------------------|
| Melhoria | Atualize TLS a versão em Tez para 1.2 |

## Amazon EMR 6.13.0 - Notas de lançamento do Tez

### Amazon EMR 6.13.0 - Mudanças no chá

| Tipo             | Descrição                                                                                                          |
|------------------|--------------------------------------------------------------------------------------------------------------------|
| Correção de bugs | REVERT <a href="#">TEZ-4295</a> : Não foi possível descompactar os dados. O comprimento do buffer é muito pequeno. |
| Correção de bugs | REVERT <a href="#">TEZ-4302</a> : NullPointerException em CodecUtils com GzipCodec                                 |



| Tipo             | Descrição                                                                                                                                        |
|------------------|--------------------------------------------------------------------------------------------------------------------------------------------------|
| Correção de bugs | REVERT <a href="#">TEZ-4234</a> : O compressor pode causar <code>IllegalArgumentExcepcion Buffer.limit</code> onde o limite excede a capacidade. |
| Correção de bugs | REVERT <a href="#">TEZ-4135</a> : Melhore a alocação de memória ao executar leituras na memória.                                                 |

## Amazon EMR 6.12.0 - Notas de lançamento do Tez

### Amazon EMR 6.12.0 - Mudanças no chá

| Tipo             | Descrição                                                                                                                                                                        |
|------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Melhoria         | Suporte adicionado para JDK 11 e JDK 17 Runtime                                                                                                                                  |
| Correção de bugs | <a href="#">TEZ-4492</a> : Atualize o Bowerrc para usar o espelho <code>bower.herokuapp</code> para evitar o problema do Bower Registry <code>_ (-2608) CERT EXPIRE BOWER</code> |
| Upgrade          | Atualizado o Surefire para 3.0.0-M7                                                                                                                                              |

## Amazon EMR 6.11.0 - Notas de lançamento do Tez

### Amazon EMR 6.11.0 - Mudanças de chá

| Tipo | Descrição                                                                                                        |
|------|------------------------------------------------------------------------------------------------------------------|
| Bug  | Corrigida a transição de estado de vértices inválida durante a limpeza em nível de vértice de dados embaralhados |

| Tipo     | Descrição                                                                                                         |
|----------|-------------------------------------------------------------------------------------------------------------------|
| Bug      | A limpeza fixa DAG ou em nível de vértice de dados aleatórios não funciona                                        |
| Melhoria | Ative <code>tez.am.dag.cleanup.on.completion</code> por padrão para limpar os dados aleatórios de concluídos DAGs |

## Amazon EMR 6.10.0 - Notas de lançamento do Tez

### Amazon EMR 6.10.0 - Mudanças no chá

| Tipo     | Descrição                                                                                                                                                   |
|----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Atributo | Habilitar <code>tez.runtime.transfer.data-via-events.enabled</code> por padrão                                                                              |
| Backport | <a href="#">TEZ-4450</a> : Corrigir falha na busca aleatória de dados quando dados aleatórios são transferidos por meio de eventos de movimentação de dados |
| Backport | <a href="#">TEZ-4460</a> : Corrige o erro de tempo limite de leitura na busca de dados aleatórios do Tez Shuffle Handler                                    |
| Backport | <a href="#">TEZ-4455</a> : Adicione um ShuffleHandler pipeline para melhorar a LoggingHandler capacidade de depuração                                       |
| Bug      | Corrija a tarefa do Tez que fica presa de forma intermitente quando a preempção da tarefa está habilitada                                                   |

## Amazon EMR 6.9.0 - Notas de lançamento do Tez

### Amazon EMR 6.9.0 - Mudanças no chá

| Tipo    | Descrição                                                                                                                        |
|---------|----------------------------------------------------------------------------------------------------------------------------------|
| Upgrade | O Tez é atualizado para 0.10.2. Para obter mais informações, consulte o <a href="#">log de alterações do Apache Tez 0.10.2</a> . |
| Upgrade | Atualizar o Hadoop para 3.3.3.                                                                                                   |
| Bug     | Desative <code>tez.runtime.transfer.data-via-events.enabled</code> por padrão devido a <a href="#">TEZ-4450</a> .                |

## Amazon EMR 6.8.0 - Notas de lançamento do Tez

### Amazon EMR 6.8.0 - Mudanças no chá

| Tipo     | Descrição                                                                                                            |
|----------|----------------------------------------------------------------------------------------------------------------------|
| Backport | <a href="#">TEZ-3363</a> : Exclua dados intermediários no nível do vértice para o Shuffle Handler                    |
| Backport | <a href="#">TEZ-4129</a> : Exclua dados de tentativas intermediárias para tentativas malsucedidas do Shuffle Handler |
| Backport | <a href="#">TEZ-4430</a> : A propriedade <code>tez.task.launch.cmd-opts</code> corrigida não funcionava              |

## Amazon EMR 6.7.0 - Notas de lançamento do Tez

### Amazon EMR 6.7.0 - Mudanças no chá

| Tipo     | Descrição                                                                           |
|----------|-------------------------------------------------------------------------------------|
| Backport | <a href="#">TEZ-4403</a> : Atualize a SLF4J versão para 1.7.36                      |
| Backport | <a href="#">TEZ-4405</a> : Substitua log4j 1.x por reload4j                         |
| Backport | <a href="#">TEZ-4411</a> : Falha na compilação do Tez: FileSaver .js não encontrado |

## Amazon EMR 6.6.0 - Notas de lançamento do Tez

### Amazon EMR 6.6.0 - Mudanças no chá

| Tipo     | Descrição                                                                                |
|----------|------------------------------------------------------------------------------------------|
| Backport | <a href="#">TEZ-3918</a> : A propriedade tez.task.log.level corrigida não funcionava.    |
| Backport | <a href="#">TEZ-4353</a> : Atualize o commons-io para 2.8.0.                             |
| Backport | <a href="#">TEZ-4114</a> : Remova a dependência direta do cais do tez.                   |
| Backport | <a href="#">TEZ-4323</a> : Os jarros Jetty foram removidos do pacote dist com -4114. TEZ |

# Apache Zeppelin

Use o Apache Zeppelin como um notebook para exploração de dados interativos. Para obter mais informações sobre o Zeppelin, consulte <https://zeppelin.apache.org/>. O Zeppelin está incluído nas EMR versões 5.0.0 e posteriores da Amazon. As versões anteriores incluem o Zeppelin como uma aplicação em sandbox. Para ter mais informações, consulte [Versões de lançamento do Amazon EMR 4.x](#).

Para acessar a interface web do Zeppelin, configure um SSH túnel para o nó principal e uma conexão proxy. Para obter mais informações, consulte [Exibir interfaces web hospedadas em EMR clusters](#).

A tabela a seguir lista a versão do Zeppelin incluída na versão mais recente da série Amazon EMR 7.x, junto com os componentes que a Amazon instala com o Zeppelin. EMR

Para a versão dos componentes instalados com o Zeppelin nesta versão, consulte Versões de componentes da [versão 7.2.0](#).

Informações sobre a versão do Zeppelin para o emr-7.2.0

| Gravadora EMR de lançamento da Amazon | Versão do Zeppelin | Componentes instalados com o 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-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 |

A tabela a seguir lista a versão do Zeppelin incluída na versão mais recente da série Amazon EMR 6.x, junto com os componentes que a Amazon instala com o Zeppelin. EMR

Para obter a versão dos componentes instalados com o Zeppelin nessa versão, consulte [Release 6.15.0 Component Versions](#).

Informações de versão do Zeppelin para o emr-6.15.0

| Gravadora EMR de lançamento da Amazon | Versão do Zeppelin | Componentes instalados com o 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 |

A tabela a seguir lista a versão do Zeppelin incluída na versão mais recente da série Amazon EMR 5.x, junto com os componentes que a Amazon instala com o Zeppelin. EMR

Para a versão dos componentes instalados com o Zeppelin nesta versão, consulte Versões de componentes da [versão 5.36.2](#).

Informações sobre a versão do Zeppelin para o emr-5.36.2

| Gravadora EMR de lançamento da Amazon | Versão do Zeppelin | Componentes instalados com o Zeppelin                |
|---------------------------------------|--------------------|------------------------------------------------------|
| emr-5.36.2                            | Zeppelin 0.10.0    | aws-sagemaker-spark-sdk, emrfs, emr-goodies, hadoop- |

| Gravadora EMR de lançamento da Amazon | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                       |
|---------------------------------------|--------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                       |                    | 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 |

## Tópicos

- [Considerações ao usar o Zeppelin na Amazon EMR](#)
- [Histórico de versões do Zeppelin](#)

## Considerações ao usar o Zeppelin na Amazon EMR

- Conecte-se ao Zeppelin usando o mesmo [método de SSH tunelamento para se conectar a outros servidores web no nó](#) principal. O servidor do Zeppelin está disponível na porta 8890.
- [As versões 5.0.0 e posteriores do Zeppelin na Amazon EMR são compatíveis com a autenticação Shiro.](#)
- O Zeppelin na Amazon EMR lança as versões 5.8.0 e posteriores com suporte ao uso do AWS Glue Data Catalog como metastore do Spark. SQL Para obter mais informações, consulte [Usando o AWS Glue Data Catalog como metastore do Spark](#). SQL
- O Zeppelin não usa algumas das configurações definidas no arquivo de `spark-defaults.conf` configuração do seu cluster, embora instrua a YARN alocar os executores dinamicamente, se você tiver definido `spark.dynamicAllocation.enabled true` Você deve definir as configurações do executor, como memória e núcleos, usando a guia Interpretar (Intérprete) do Zeppelin e reiniciar o intérprete para que eles sejam usados.

- As EMR versões 6.10.0 e superiores da Amazon oferecem suporte à integração do Apache Zeppelin com o Apache Flink. Consulte [Usar trabalhos do Flink pelo Zeppelin no Amazon EMR](#) Para mais informações.
- O Zeppelin na Amazon não EMR suporta o intérprete SparkR.

## Histórico de versões do Zeppelin

A tabela a seguir lista a versão do Zeppelin incluída em cada versão de lançamento da AmazonEMR, junto com os componentes instalados com o aplicativo. Para obter as versões dos componentes em cada versão, consulte a seção Versão de componentes da sua versão em [Versões de lançamento do Amazon EMR 7.x](#), [Versões de lançamento do Amazon EMR 6.x](#) ou [Versões de lançamento do Amazon EMR 5.x](#).

### Informações da versão do Zeppelin

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                                 |
|------------------------------|--------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-7.2.0                    | 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 |
| 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-                                                                                                                                                                                                                  |



| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o 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                                                                                                            |
| emr-7.1.0                    | 0.10.1             | 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, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                                                          |
|------------------------------|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-7.0.0                    | 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                          |
| emr-6.15.0                   | 0.10.1             | 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, hudi, hudi-spark, livy-server, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                                                           |
|------------------------------|--------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.14.0                   | 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 |
| emr-6.13.0                   | 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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                                                           |
|------------------------------|--------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.12.0                   | 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 |
| emr-6.11.1                   | 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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                                                           |
|------------------------------|--------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.11.0                   | 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 |
| emr-6.10.1                   | 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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                                                           |
|------------------------------|--------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.10.0                   | 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 |
| emr-6.9.1                    | 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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                                                           |
|------------------------------|--------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.9.0                    | 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 |
| emr-6.8.1                    | 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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                                                           |
|------------------------------|--------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.8.0                    | 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 |
| 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 |



| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                                                           |
|------------------------------|--------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                                                           |
|------------------------------|--------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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-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, zeppelin-server                   |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                                        |
|------------------------------|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                                        |
|------------------------------|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                                        |
|------------------------------|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                                        |
|------------------------------|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                                        |
|------------------------------|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                                        |
|------------------------------|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |



| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                                        |
|------------------------------|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                                        |
|------------------------------|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                                        |
|------------------------------|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                                        |
|------------------------------|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                                        |
|------------------------------|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                                        |
|------------------------------|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                                        |
|------------------------------|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                                        |
|------------------------------|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |



| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                                        |
|------------------------------|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                                        |
|------------------------------|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                           |
|------------------------------|--------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                           |
|------------------------------|--------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                           |
|------------------------------|--------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                           |
|------------------------------|--------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                            |
|------------------------------|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-httpfs-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-5.17.2                   | 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, r, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                           |
|------------------------------|--------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |



| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                           |
|------------------------------|--------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                           |
|------------------------------|--------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                           |
|------------------------------|--------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                           |
|------------------------------|--------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                           |
|------------------------------|--------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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    |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                        |
|------------------------------|--------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                         |
|------------------------------|--------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-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-5.11.4                   | 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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                        |
|------------------------------|--------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |



| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                                                        |
|------------------------------|--------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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-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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                               |
|------------------------------|--------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.10.1                   | 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-5.10.0                   | 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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                               |
|------------------------------|--------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                               |
|------------------------------|--------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| emr-5.8.2                    | 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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                               |
|------------------------------|--------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.8.1                    | 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                               |
|------------------------------|--------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                                               |
|------------------------------|--------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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 |
| 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 |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                   |
|------------------------------|--------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.5.4                    | 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-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| 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-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |



| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                  |
|------------------------------|--------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.5.2                    | 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 |
| emr-5.5.1                    | 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, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                  |
|------------------------------|--------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.5.0                    | 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, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| emr-5.4.1                    | 0.7.0              | 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, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                   |
|------------------------------|--------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| 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-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                   |
|------------------------------|--------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| emr-5.3.0                    | 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-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                  |
|------------------------------|--------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| emr-5.2.2                    | 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-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                   |
|------------------------------|--------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| emr-5.2.0                    | 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-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                   |
|------------------------------|--------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.1.1                    | 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-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| 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-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |

| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                   |
|------------------------------|--------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| 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-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |



| Gravadora Amazon EMR Release | Versão do Zeppelin | Componentes instalados com o Zeppelin                                                                                                                                                                                                                                                   |
|------------------------------|--------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 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-resourcemanager, spark-client, spark-history-server, spark-on-yarn, spark-yarn-slave, zeppelin-server |
| 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

O Apache ZooKeeper é um serviço centralizado para manter informações de configuração, nomear, fornecer sincronização distribuída e fornecer serviços de grupo. Para obter mais informações sobre ZooKeeper, consulte <http://zookeeper.apache.org/>.

A tabela a seguir lista a versão ZooKeeper incluída na versão mais recente da série Amazon EMR 7.x, junto com os componentes com os quais a Amazon é EMR instalada. ZooKeeper

Para a versão dos componentes instalados ZooKeeper nesta versão, consulte [Versões de componentes da versão 7.2.0](#).

ZooKeeper informações sobre a versão do emr-7.2.0

| Gravadora EMR de lançamento da Amazon | ZooKeeper Versão | Componentes instalados com 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 |

A tabela a seguir lista a versão ZooKeeper incluída na versão mais recente da série EMR 6.x da Amazon, junto com os componentes com os quais a Amazon é EMR instalada. ZooKeeper

Para a versão dos componentes instalados ZooKeeper nesta versão, consulte [Versões de componentes da versão 6.15.0](#).

## ZooKeeper informações da versão do emr-6.15.0

| Gravadora EMR de lançamento da Amazon | ZooKeeper Versão | Componentes instalados com ZooKeeper                                                                                                                                                                                                                              |
|---------------------------------------|------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-6.15.0                            | ZooKeeper 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 |

A tabela a seguir lista a versão ZooKeeper incluída na versão mais recente da série Amazon EMR 5.x, junto com os componentes com os quais a Amazon é EMR instalada. ZooKeeper

Para a versão dos componentes instalados ZooKeeper nesta versão, consulte Versões de componentes da [versão 5.36.2](#).

## ZooKeeper informações da versão do emr-5.36.2

| Gravadora EMR de lançamento da Amazon | ZooKeeper Versão | Componentes instalados com ZooKeeper                                                                                                                                                                                                                              |
|---------------------------------------|------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.36.2                            | ZooKeeper 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 |

## Tópicos

- [ZooKeeper histórico de lançamentos](#)

## ZooKeeper histórico de lançamentos

A tabela a seguir lista a versão ZooKeeper incluída em cada versão de lançamento da AmazonEMR, junto com os componentes instalados com o aplicativo. Para obter as versões dos componentes em cada versão, consulte a seção Versão de componentes da sua versão em [Versões de lançamento do Amazon EMR 7.x](#), [Versões de lançamento do Amazon EMR 6.x](#) ou [Versões de lançamento do Amazon EMR 5.x](#).

### ZooKeeper informações da versão

| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com 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 |

| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com ZooKeeper                                                                                                                                                                                                                              |
|------------------------------|------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-7.1.0                    | 3.9.1            | 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, 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-https-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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com 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 |

| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com 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 |

| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com 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 |



| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com 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 |

| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com 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 |

| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com 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 |

| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com 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 |

| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com 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 |

| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com 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 |



| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com 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 |

| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com ZooKeeper                                                                                                                                                                                                                              |
|------------------------------|------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-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 |

| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com 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 |

| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com 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 |

| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com 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 |

| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com 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 |

| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com ZooKeeper                                                                                                                                                                                                                              |
|------------------------------|------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.20.1                   | 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-5.20.0                   | 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-5.19.1                   | 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 |

| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com 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 |



| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com 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 |

| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com 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 |

| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com 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 |

| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com 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 |

| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, hadoop-yarn-timeline-server, zookeeper-client, zookeeper-server |

| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com 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 |

| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com 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 |

| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com 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 |



| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com 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 |

| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com 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 |

| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com 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 |

| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com 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 |

| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com ZooKeeper                                                                                                                                                                                                  |
|------------------------------|------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| emr-5.5.1                    | 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, 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-https-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-https-server, hadoop-kms-server, hadoop-yarn-nodemanager, hadoop-yarn-resource-manager, zookeeper-client, zookeeper-server |

| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, zookeeper-client, zookeeper-server |

| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com 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-resourcemanager, 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-resourcemanager, 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-resourcemanager, zookeeper-client, zookeeper-server |

| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com 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 |



| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com 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 |

| Gravadora Amazon EMR Release | ZooKeeper Versão | Componentes instalados com 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 |

# Conectores e utilitários

EMR Amazon fornece vários conectores e utilitários para acessar outros AWS serviços como fontes de dados. Você pode geralmente acessar os dados nesses serviços dentro de um programa. Por exemplo, você pode especificar um stream do Kinesis em uma consulta, script do Pig ou MapReduce aplicativo do Hive e, em seguida, operar com esses dados.

## Tópicos

- [Exporte, importe, consulte e junte tabelas no DynamoDB usando a Amazon EMR](#)
- [Kinesis](#)
- [S3 DistCp \(s3-dist-cp\)](#)
- [Limpendo após falhas em trabalhos do S3 DistCp](#)

## Exporte, importe, consulte e junte tabelas no DynamoDB usando a Amazon EMR

### Note

O Amazon EMR -DynamoDB Connector é de código aberto em GitHub. Para obter mais informações, consulte <https://github.com/aws-labs/emr-dynamodb-connector>.


O DynamoDB é um serviço SQL sem banco de dados totalmente gerenciado que fornece desempenho rápido e previsível com escalabilidade perfeita. Os desenvolvedores podem criar uma tabela de banco de dados e ampliar seu tráfego de solicitação ou armazenamento sem limites. O DynamoDB distribui automaticamente dados e tráfego para a tabela através de um número suficiente de servidores, a fim de controlar a capacidade de solicitação especificada pelo cliente e a quantidade de dados armazenados, enquanto mantém um desempenho consistente e rápido. Usando a Amazon EMR e o Hive, você pode processar com rapidez e eficiência grandes quantidades de dados, como dados armazenados no DynamoDB. Para obter mais informações sobre o DynamoDB, consulte o [Guia do desenvolvedor do Amazon DynamoDB](#).

O Apache Hive é uma camada de software que você pode usar para consultar, mapear e reduzir clusters usando uma linguagem de consulta simplificada chamada SQL HiveQL. Ele é executado sobre a arquitetura do Hadoop. Para obter mais informações sobre o Hive e o HiveQL, acesse o

documento [HiveQL Language Manual](#). Para obter mais informações sobre o Hive e a AmazonEMR, consulte [Apache Hive](#).

Você pode usar a Amazon EMR com uma versão personalizada do Hive que inclui conectividade com o DynamoDB para realizar operações nos dados armazenados no DynamoDB:

- Carregar dados do DynamoDB no Hadoop Distributed File System HDFS () e usá-los como entrada em um cluster da Amazon. EMR
- Consultar dados dinâmicos do DynamoDB usando instruções do SQL tipo -like (HiveQL).
- Unir dados armazenados no DynamoDB e exportar ou consultar com base nos dados unidos.
- Exportar dados armazenados no DynamoDB para o Amazon S3.
- Importar dados armazenados no Amazon S3 para o DynamoDB.


 Note

[O Amazon EMR -DynamoDB Connector não oferece suporte a clusters configurados para usar a autenticação Kerberos.](#)

Para realizar cada uma das tarefas a seguir, você iniciará um EMR cluster da Amazon, especificará a localização dos dados no DynamoDB e emitirá comandos do Hive para manipular os dados no DynamoDB.

Há várias maneiras de iniciar um EMR cluster da Amazon: você pode usar o EMR console da Amazon, a interface de linha de comando (CLI) ou você pode programar seu cluster usando um AWS SDK ou o Amazon EMRAPI. Você também pode escolher se deseja executar um cluster do Hive interativamente ou a partir de um script. Nesta seção, mostraremos como iniciar um cluster interativo do Hive a partir do EMR console da Amazon e do CLI.

Usar o Hive, interativamente, é uma ótima maneira de testar o desempenho das consultas e ajustar seu aplicativo. Depois de estabelecer um conjunto de comandos do Hive que serão executados regularmente, considere criar um script do Hive que a Amazon EMR possa executar para você.

 Warning

As operações de EMR leitura ou gravação da Amazon em uma tabela do DynamoDB são contabilizadas em relação à taxa de transferência provisionada estabelecida, aumentando potencialmente a frequência das exceções da taxa de transferência provisionada. Para

solicitações grandes, a Amazon EMR implementa novas tentativas com recuo exponencial para gerenciar a carga de solicitações na tabela do DynamoDB. Executar EMR trabalhos da Amazon simultaneamente com outros tráfegos pode fazer com que você exceda o nível de taxa de transferência provisionado alocado. Você pode monitorar isso verificando a `ThrottleRequests` métrica na Amazon CloudWatch. Se a carga da solicitação for muito alta, você poderá reiniciar o cluster e definir o [Configuração de porcentagem de leitura](#) ou [Configuração de porcentagem de gravação](#) para um valor menor para acelerar as operações da Amazon. Para obter mais informações sobre as configurações de throughput do DynamoDB, consulte [Provisioned throughput](#).

Se uma tabela estiver configurada para o [Modo sob demanda](#), será necessário alterá-la novamente para o modo provisionado antes de executar uma operação de exportação ou importação. Os dutos precisam de uma taxa de transferência para calcular os recursos a serem usados a partir de um `DynamoDBTable`. O modo sob demanda remove o throughput provisionado. Para provisionar a capacidade de transferência, você pode usar as métricas do Amazon CloudWatch Events para avaliar a taxa de transferência agregada usada por uma tabela.

## Tópicos

- [Configurar uma tabela do Hive para executar comandos do Hive](#)
- [Exemplos de comandos do Hive para exportar, importar e consultar dados no DynamoDB](#)
- [Otimizando o desempenho das EMR operações da Amazon no DynamoDB](#)

## Configurar uma tabela do Hive para executar comandos do Hive

O Apache Hive é um aplicativo de armazém de dados que você pode usar para consultar dados contidos nos EMR clusters da Amazon usando uma linguagem SQL semelhante. Para obter mais informações sobre o Hive, consulte <http://hive.apache.org/>.

O procedimento a seguir pressupõe que você já tenha criado um cluster e especificado um par de EC2 chaves da Amazon. Para saber como começar a criar clusters, consulte [Introdução à Amazon EMR](#) no Guia de EMR gerenciamento da Amazon.

## Configurar o Hive para usar MapReduce

Quando você usa o Hive na Amazon EMR para consultar tabelas do DynamoDB, podem ocorrer erros se o Hive usar o mecanismo de execução padrão, o Tez. Por esse motivo, ao criar um cluster

com o Hive que se integre ao DynamoDB conforme descrito nesta seção, recomendamos que você use uma classificação de configuração que defina o Hive a ser usado. MapReduce Para obter mais informações, consulte [Configurar aplicações](#).

O trecho a seguir mostra a classificação de configuração e a propriedade a serem usadas para definir MapReduce como mecanismo de execução do Hive:

```
[
 {
 "Classification": "hive-site",
 "Properties": {
 "hive.execution.engine": "mr"
 }
 }
]
```

Para executar comandos do Hive interativamente

1. Conecte-se ao nó principal. Para obter mais informações, consulte [Conecte-se ao nó principal usando SSH](#) o Amazon EMR Management Guide.
2. No prompt de comando do nó principal atual, digite `hive`.

Você verá um prompt do Hive: `hive>`

3. Insira um comando do Hive que mapeie uma tabela na aplicação Hive para os dados no DynamoDB. Essa tabela serve de referência para os dados armazenados no Amazon DynamoDB; os dados não são armazenados localmente no Hive, e qualquer consulta usando essa tabela é executada com base nos dados dinâmicos no DynamoDB, consumindo a capacidade de leitura ou gravação da tabela sempre que um comando é executado. Se você pretende executar vários comandos do Hive no mesmo conjunto de dados, considere exportá-lo primeiro.

O exemplo a seguir mostra a sintaxe para mapear uma tabela do Hive para uma tabela do 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..."
```

Ao criar uma tabela no Hive pelo DynamoDB, você deve criá-la como uma tabela externa usando a palavra-chave `EXTERNAL`. A diferença entre tabelas externas e internas é que os dados nas tabelas internas são excluídos quando uma tabela interna é descartada. Este não é o comportamento desejado no caso de uma conexão com o Amazon DynamoDB e, portanto, somente tabelas externas são compatíveis.

Por exemplo, o seguinte comando do Hive cria uma tabela chamada `hivetable1` no Hive que referencia a tabela do DynamoDB chamada `dynamodhtable1`. A tabela `dynamodhtable1` do DynamoDB tem um esquema de chave primária. `hash-and-range` O elemento da chave de hash é `name` (do tipo `string`), o elemento de chave de intervalo é `year` (do tipo numérico), e cada item possui um valor de atributo para `holidays` (do tipo conjunto de strings).

```
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");
```

A linha 1 usa a instrução HiveQL `CREATE EXTERNAL TABLE`. Para `hivetable1`, você precisa estabelecer uma coluna para cada par de nome-valor de atributo na tabela do DynamoDB e fornecer o tipo de dados. Esses valores não diferenciam maiúsculas de minúsculas, e você pode dar qualquer nome (exceto palavras reservadas) para as colunas.

A linha 2 usa a instrução `STORED BY`. O valor de `STORED BY` é o nome da classe que manipula a conexão entre o Hive e o DynamoDB. Ele deve ser definido como `'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'`.

A linha 3 usa a instrução `TBLPROPERTIES` para associar `"hivetable1"` com a tabela e o esquema corretos no DynamoDB. Forneça a `TBLPROPERTIES` valores para os parâmetros `dynamodb.table.name` e `dynamodb.column.mapping`. Esses valores diferenciam maiúsculas de minúsculas.

**Note**

Todos os nomes de atributo do DynamoDB para a tabela devem ter colunas correspondentes na tabela do Hive. Dependendo da sua EMR versão da Amazon, os seguintes cenários ocorrerão se o one-to-one mapeamento não existir:

- Na Amazon EMR versão 5.27.0 e posterior, o conector tem validações que garantem um mapeamento one-to-one entre nomes de atributos e colunas do DynamoDB na tabela do Hive. Ocorrerá um erro se o one-to-one mapeamento não existir.
- Na Amazon EMR versão 5.26.0 e anteriores, a tabela do Hive não conterá o par nome-valor do DynamoDB. Se você não mapear os atributos de chave primária do DynamoDB, o Hive gerará um erro. Se você não mapear um atributo de chave não primária, nenhum erro será gerado, mas os dados não serão vistos na tabela do Hive. Se os tipos de dados não corresponderem, o valor será nulo.

Em seguida, você poderá começar a executar operações do Hive em `hivetable1`. As consultas executadas em `hivetable1` são internamente executadas na tabela do DynamoDB `dynamodhtable1` da sua conta do DynamoDB, consumindo unidades de leitura ou gravação com cada execução.

Quando você executa consultas do Hive em uma tabela do DynamoDB, precisa certificar-se de ter provisionado uma quantidade suficiente de unidades de capacidade de leitura.

Por exemplo, suponha que você tenha provisionado 100 unidades de capacidade de leitura para a sua tabela do DynamoDB. Isso permitirá que você realize 100 leituras, ou 409.600 bytes, por segundo. Se essa tabela contiver 20 GB de dados (21.474.836.480 bytes) e sua consulta do Hive realizar uma verificação de tabela completa, você poderá estimar quanto tempo a consulta demorará para ser executada:

$$21.474.836.480/409.600 = 52.429 \text{ segundos} = 14,56 \text{ horas}$$

A única maneira de diminuir o tempo necessário seria ajustar as unidades de capacidade de leitura na tabela DynamoDB de origem. Adicionar mais EMR nós da Amazon não ajudará.

Na saída do Hive, a porcentagem de conclusão é atualizada quando um ou mais processos de mapeador são finalizados. Para uma tabela grande do DynamoDB com uma configuração baixa de capacidade de leitura provisionada, a saída de percentual de conclusão pode não ser atualizada por um longo tempo. No caso acima, o trabalho parecerá estar 0% concluído por várias horas.



Para obter um status mais detalhado sobre o progresso do seu trabalho, acesse o EMR console da Amazon; você poderá visualizar o status individual da tarefa do mapeador e as estatísticas das leituras de dados. Você também pode fazer logon na interface do Hadoop no nó principal e visualizar as estatísticas do Hadoop. Isso mostrará o status de tarefas map individuais e algumas estatísticas de leitura de dados. Para obter mais informações, consulte os tópicos a seguir.

- [Interfaces Web hospedadas no nó principal](#)
- [Visualizar as interfaces Web do Hadoop](#)

Para obter mais informações sobre instruções HiveQL de exemplo para realizar tarefas, como exportar ou importar dados do DynamoDB e unir tabelas, consulte [Exemplos de comandos do Hive para exportar, importar e consultar dados no DynamoDB](#).

Para cancelar uma solicitação do Hive

Quando você executa uma consulta do Hive, a resposta inicial do servidor inclui o comando para cancelar a solicitação. Para cancelar a solicitação a qualquer momento no processo, use Kill Command (Comando Kill) na resposta do servidor.

1. Insira `Ctrl+C` para sair do cliente de linha de comando.
2. No prompt do shell, insira o Kill Command (Comando Kill) da resposta do servidor inicial à sua solicitação.

Como alternativa, você pode executar o seguinte comando na linha de comando do nó principal para eliminar o trabalho do Hadoop, onde *job-id* é o identificador da tarefa do Hadoop e pode ser recuperado da interface de usuário do Hadoop.

```
hadoop job -kill job-id
```

## Tipos de dados para o Hive e o DynamoDB

A tabela a seguir mostra os tipos de dados do Hive disponíveis, o tipo do DynamoDB padrão ao qual eles correspondem e os tipos do DynamoDB alternativos para os quais eles também podem mapear.

| Tipo do Hive         | Tipo do DynamoDB padrão | Tipos alternativos do DynamoDB                                                  |
|----------------------|-------------------------|---------------------------------------------------------------------------------|
| string               | string (S)              |                                                                                 |
| bigint ou duplo      | número (N)              |                                                                                 |
| binary               | binário (B)             |                                                                                 |
| boolean              | booleano () BOOL        |                                                                                 |
| array                | lista (L)               | conjunto de números (NS), conjunto de strings (SS) ou conjunto de binários (BS) |
| mapa<string, string> | item                    | mapa (M)                                                                        |
| mapa<string,?>       | mapa (M)                |                                                                                 |
|                      | nulo () NULL            |                                                                                 |

Se você deseja gravar seus dados do Hive como um tipo alternativo correspondente do DynamoDB, ou se os dados do DynamoDB contiverem valores de atributo de um tipo alternativo do DynamoDB, você poderá especificar a coluna e o tipo do DynamoDB usando o parâmetro `dynamodb.type.mapping`. O exemplo a seguir mostra a sintaxe para especificar um tipo alternativo de mapeamento.

```
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");
```

O parâmetro de mapeamento de tipo é opcional e só tem que ser especificado para as colunas que usam tipos alternativos.

Por exemplo, o comando do Hive a seguir cria uma tabela chamada `hivetable2` que referencia tabela `dynamodbtbl2` do DynamoDB. Ela é semelhante à `hivetable1`, exceto pelo fato de mapear a coluna `col3` para o tipo de conjunto de strings (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");
```

No Hive, `hivetable1` e `hivetable2` são idênticas. No entanto, quando os dados dessas tabelas forem gravados em suas tabelas correspondentes do DynamoDB, a `dynamodbtbl1` conterá listas, enquanto a `dynamodbtbl2` conterá conjuntos de strings.

Se você quiser gravar valores `null` do Hive como atributos do tipo `null` do DynamoDB, poderá fazer isso usando o parâmetro `dynamodb.null.serialization`. O exemplo a seguir mostra a sintaxe para especificar a serialização `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");
```

O parâmetro de serialização nula é opcional e será definido como `false` se não especificado. Os atributos `null` do DynamoDB são lidos como valores `null` no Hive, independentemente da configuração do parâmetro. As coleções do Hive com valores `null` só poderão ser gravadas no DynamoDB se o parâmetro de serialização nula for especificado como `true`. Caso contrário, ocorrerá um erro do Hive.

O tipo `bigint` no Hive equivale ao tipo longo no Java, enquanto o tipo duplo no Hive equivale ao tipo duplo no Java em termos de precisão. Isso significa que, se você tiver dados numéricos armazenados no DynamoDB cuja precisão seja maior que a disponível nos tipos de dados do Hive, usar o Hive para exportar, importar ou referenciar os dados do DynamoDB pode resultar na perda de precisão ou em uma falha da consulta do Hive.

Exportações do tipo binário do DynamoDB para o Amazon Simple Storage Service HDFS (Amazon S3) ou são armazenadas como uma string codificada em Base64. Se você estiver importando dados

do Amazon S3 HDFS ou para o tipo binário do DynamoDB, eles devem ser codificados como uma string Base64.

## Opções do Hive

Você pode definir as seguintes opções do Hive para gerenciar a transferência de dados do Amazon DynamoDB. Essas opções só persistem na sessão atual do Hive. Se você fechar o prompt de comando do Hive e reabri-lo mais tarde no cluster, essas configurações retornarão aos valores padrão.

| Opções do Hive                                | Descrição                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|-----------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>dynamodb.throughput.read.percent</code> | <p>Defina a taxa de operações de leitura para manter sua taxa de throughput provisionado do DynamoDB no intervalo alocado para sua tabela. O valor está entre 0.1 e 1.5, inclusive.</p> <p>O valor de 0,5 é a taxa de leitura padrão, que significa que o Hive tentará consumir metade dos recursos de throughput provisionado de leitura na tabela. Aumentar esse valor acima de 0,5 aumenta a taxa de solicitações de leitura. Diminuí-lo abaixo de 0,5 diminui a taxa de solicitações de leitura. Essa taxa de leituras é aproximada. A taxa de leituras real dependerá de fatores como, por exemplo, a existência de distribuição uniforme de chaves no DynamoDB.</p> <p>Se você perceber que o seu throughput provisionado está sendo frequentemente excedido pela operação do Hive, ou se o tráfego de leituras dinâmicas estiver muito limitado, reduza esse valor abaixo de 0.5.</p> <p>Se você tiver capacidade suficiente e quiser uma operação do Hive mais rápida, defina esse valor acima de 0.5. Também é possível configurar o excesso de assinaturas, definindo o valor como até 1.5 se você acredita que existem operações de entrada/saída não utilizadas disponíveis.</p> |

| Opções do Hive                                 | Descrição                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>dynamodb.throughput.write.percent</code> | <p>Defina a taxa de operações de gravação para manter sua taxa de throughput provisionado do DynamoDB no intervalo alocado para sua tabela. O valor está entre 0.1 e 1.5, inclusive.</p> <p>O valor de 0,5 é a taxa de gravação padrão, que significa que o Hive tentará consumir metade dos recursos de throughput provisionado de gravação na tabela. Aumentar esse valor acima de 0,5 aumenta a taxa de solicitações de gravação. Diminuí-lo abaixo de 0,5 diminui a taxa de solicitações de gravação. Essa taxa de gravações é aproximada. A taxa de gravações real dependerá de fatores como se há uma distribuição uniforme de chaves no DynamoDB</p> <p>Se você perceber que o seu throughput provisionado está sendo frequentemente excedido pela operação do Hive, ou se o tráfego de gravações dinâmicas estiver muito limitado, reduza esse valor abaixo de 0.5. Se você tiver capacidade suficiente e quiser uma operação do Hive mais rápida, defina esse valor acima de 0.5. Também é possível configurar o excesso de assinaturas, definindo o valor como até 1.5 se você acredita que existem operações de entrada/saída não utilizadas disponíveis ou se este for o upload de dados inicial na tabela e ainda não houver tráfego dinâmico.</p> |
| <code>dynamodb.endpoint</code>                 | <p>Especifique o endpoint para o serviço do DynamoDB. Para obter mais informações sobre os endpoints do DynamoDB disponíveis, consulte <a href="#">Regions and endpoints</a>.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| <code>dynamodb.max.map.tasks</code>            | <p>Especifique o número máximo de tarefas map ao ler dados do DynamoDB. Esse valor deve ser igual ou maior que 1.</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |

| Opções do Hive                       | Descrição                                                                                                                                                                                                                                  |
|--------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>dynamodb.retry.duration</code> | Especifique o número de minutos a serem usados como a duração do tempo limite para tentar novamente os comandos do Hive. Esse valor deve ser um número inteiro igual a ou maior que 0. A duração do tempo limite padrão é de dois minutos. |

Essas opções são definidas usando o comando SET, como mostra o exemplo a seguir.

```
SET dynamodb.throughput.read.percent=1.0;

INSERT OVERWRITE TABLE s3_export SELECT *
FROM hiveTableName;
```

## Exemplos de comandos do Hive para exportar, importar e consultar dados no DynamoDB

Os exemplos a seguir usam comandos do Hive para realizar operações como exportar dados para o Amazon S3 HDFS ou importar dados para o DynamoDB, unir tabelas, consultar tabelas e muito mais.

As operações em uma tabela do Hive referenciam dados armazenados no DynamoDB. Os comandos do Hive estão sujeitos às configurações de throughput provisionado da tabela do DynamoDB, e os dados recuperados incluem os dados gravados na tabela do DynamoDB na ocasião em que a solicitação de operação do Hive é processada pelo DynamoDB. Se o processo de recuperação de dados demorar muito, alguns dados retornados pelo comando do Hive podem ter sido atualizados no DynamoDB desde que o comando do Hive foi iniciado.

Os comandos do Hive DROP TABLE e CREATE TABLE atuam apenas nas tabelas locais do Hive e não criam nem descartam tabelas no DynamoDB. Se a sua consulta do Hive fizer referência a uma tabela no DynamoDB, essa tabela já deverá existir antes da execução da consulta. Para obter mais informações sobre como criar e excluir tabelas no Amazon DynamoDB, consulte [Working with tables in DynamoDB](#) no Guia do desenvolvedor do Amazon DynamoDB.

**Note**

Quando você mapear uma tabela do Hive para um local no Amazon S3, não a mapeie para o caminho raiz do bucket, `s3://mybucket`, pois isso poderá causar erros quando o Hive gravar os dados no Amazon S3. Em vez disso, mapeie a tabela para um subcaminho do bucket, `s3://mybucket/mypath`.

## Exportar dados do DynamoDB

Você pode usar o Hive para exportar dados do DynamoDB.

Exportar uma tabela do DynamoDB para um bucket do Amazon S3

- Crie uma tabela do Hive que faça referência aos dados armazenados no DynamoDB. Em seguida, você pode chamar o `INSERT OVERWRITE` comando para gravar os dados em um diretório externo. No exemplo a seguir, `s3://bucketname/path/subpath/` é um caminho válido no Amazon S3. Ajuste as colunas e os tipos de dados no `CREATE` comando para que correspondam aos valores no seu DynamoDB. Você pode usar isso para criar um arquivamento dos seus dados do DynamoDB no 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" = "dynamodbtable1",
"dynamodb.column.mapping" = "col1:name,col2:year,col3:holidays");

INSERT OVERWRITE DIRECTORY 's3://bucketname/path/subpath/' SELECT *
FROM hiveTableName;
```

Exportar uma tabela do DynamoDB para um bucket do Amazon S3 usando formatação

- Crie uma tabela externa que referencie um local no Amazon S3. Isso é mostrado abaixo como `s3_export`. Durante a `CREATE` chamada, especifique a formatação da linha para a tabela. Então, quando você usa `INSERT OVERWRITE` para exportar dados do DynamoDB para `s3_export`, os dados são gravados no formato especificado. No exemplo a seguir, os dados são gravados como valores separados por vírgula (,)CSV.

```
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");

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;
```

Exportar uma tabela do DynamoDB para um bucket do Amazon S3 sem especificar um mapeamento de colunas

- Crie uma tabela do Hive que faça referência aos dados armazenados no DynamoDB. Isso é semelhante ao exemplo anterior, com a diferença de que você não está especificando um mapeamento de colunas. A tabela deve ter exatamente uma coluna do tipo `map<string, string>`. Se você criar uma tabela EXTERNAL no Amazon S3, poderá chamar o comando `INSERT OVERWRITE` para gravar os dados do DynamoDB para o Amazon S3. Você pode usar isso para criar um arquivamento dos seus dados do DynamoDB no Amazon S3. Como não há mapeamento de colunas, você não pode consultar tabelas exportadas dessa maneira. A exportação de dados sem especificar um mapeamento de coluna está disponível no Hive 0.8.1.5 ou posterior, que é compatível com o Amazon 2.2. EMR AMI x e mais tarde.

```
CREATE EXTERNAL TABLE hiveTableName (item map<string,string>)
STORED BY 'org.apache.hadoop.hive.dynamodb.DynamoDBStorageHandler'
TBLPROPERTIES ("dynamodb.table.name" = "dynamodhtable1");

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;
```



Exportar uma tabela do DynamoDB para um bucket do Amazon S3 usando a compactação de dados

- O Hive oferece vários codecs de codec, que você pode definir durante a sua sessão do Hive. Isso faz com que os dados exportados sejam compactados no formato especificado. O exemplo a seguir compacta os arquivos exportados usando o algoritmo 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" = "dynamodbtable1",
"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;
```

Os codecs de compactação disponíveis são:

- 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

## Para exportar uma tabela do DynamoDB para HDFS

- Use o seguinte comando do Hive, onde *hdfs:///directoryName* é um HDFS caminho válido e *hiveTableName* é uma tabela no Hive que faz referência ao DynamoDB. Essa operação de exportação é mais rápida do que exportar uma tabela do DynamoDB para o Amazon S3 porque o Hive 0.7.1.1 é HDFS usado como uma etapa intermediária ao exportar dados para o Amazon S3. O exemplo a seguir também mostra como definir `dynamodb.throughput.read.percent` como 1.0 a fim de aumentar a taxa de solicitação de leitura.

```
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");

SET dynamodb.throughput.read.percent=1.0;

INSERT OVERWRITE DIRECTORY 'hdfs:///directoryName' SELECT * FROM hiveTableName;
```

Você também pode exportar dados HDFS usando a formatação e a compactação, conforme mostrado acima, para a exportação para o Amazon S3. Para fazer isso, basta substituir o diretório Amazon S3 nos exemplos acima por um HDFS diretório.

## Para ler dados não imprimíveis de UTF -8 caracteres no Hive

- Você pode ler e gravar dados não imprimíveis de UTF -8 caracteres com o Hive usando a `STORED AS SEQUENCEFILE` cláusula ao criar a tabela. A SequenceFile é o formato de arquivo binário do Hadoop; você precisa usar o Hadoop para ler esse arquivo. O exemplo a seguir mostra como exportar dados do DynamoDB para o Amazon S3. Você pode usar essa funcionalidade para lidar com caracteres codificados em UTF -8 não imprimíveis.

```
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");

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;
```

## Importar dados para o DynamoDB

Ao gravar dados no DynamoDB usando o Hive, você deve se certificar de que o número de unidades de capacidade de gravação seja maior do que o número de mapeadores no cluster. Por exemplo, clusters executados em instâncias EC2 m1.xlarge produzem 8 mapeadores por instância. No caso de um cluster que tem 10 instâncias, isso significaria um total de 80 mapeadores. Se as suas unidades de capacidade de gravação não forem maiores que o número de mapeadores no cluster, a operação de gravação do Hive poderá consumir todo o throughput de gravação ou tentar consumir mais throughput do que o provisionado. Para obter mais informações sobre o número de mapeadores produzidos por cada tipo de EC2 instância, consulte [Configurar o Hadoop](#).

O número de mapeadores no Hadoop é controlado pelas divisões de entradas. Se houver poucas divisões, seu comando de gravação talvez não seja capaz de consumir todo o throughput de gravação disponível.

Se um item com a mesma chave existir na tabela do DynamoDB de destino, ele será substituído. Se houver nenhum item com a chave na tabela do DynamoDB de destino, o item será inserido.

### Importar uma tabela do Amazon S3 para o DynamoDB

- Você pode usar a Amazon EMR (AmazonEMR) e o Hive para gravar dados do Amazon S3 no 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;
```

## Importar uma tabela de um bucket do Amazon S3 para o DynamoDB sem especificar um mapeamento de colunas

- Crie uma tabela EXTERNAL que faça referência a dados armazenados no Amazon S3 que foram exportados anteriormente do DynamoDB. Antes de importar, verifique se a tabela exista no DynamoDB e tem o mesmo esquema de chave da tabela do DynamoDB exportada anteriormente. Além disso, a tabela deve ter exatamente uma coluna, do tipo `map<string, string>`. Se você criar uma tabela do Hive vinculada ao DynamoDB, poderá chamar o comando `INSERT OVERWRITE` para gravar os dados do Amazon S3 no DynamoDB. Como não há mapeamento de colunas, você não pode consultar tabelas importadas dessa maneira. A importação de dados sem especificar um mapeamento de coluna está disponível no Hive 0.8.1.5 ou posterior, que é compatível com o Amazon 2.2.3 e versões posteriores. EMR AMI

```
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" = "dynamoddbtable1");

INSERT OVERWRITE TABLE hiveTableName SELECT *
FROM s3TableName;
```

## Para importar uma tabela do HDFS DynamoDB

- Você pode usar a Amazon EMR e o Hive para gravar dados no 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;
```

## Consultar dados no DynamoDB

Os exemplos a seguir mostram as várias maneiras pelas quais você pode usar EMR a Amazon para consultar dados armazenados no DynamoDB.

Para encontrar o maior valor para uma coluna mapeada (**max**)

- Use os comandos do Hive como os seguintes. No primeiro comando, a CREATE instrução cria uma tabela do Hive que faz referência aos dados armazenados no DynamoDB. Em seguida, a SELECT instrução usa essa tabela para consultar dados armazenados no DynamoDB. O exemplo a seguir localiza o maior pedido efetuado por um determinado cliente.

```
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;
```

Para agregar dados usando a cláusula **GROUP BY**

- Você pode usar a cláusula GROUP BY para coletar dados em vários registros. Muitas vezes, isso é usado com uma função agregada, como sum, count, min ou max. O exemplo a seguir gera uma lista dos maiores pedidos de clientes que efetuaram mais de três pedidos.

```
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;

```

## Unir duas tabelas do DynamoDB

- O exemplo a seguir mapeia duas tabelas do Hive para dados armazenados no DynamoDB. Em seguida, ele chama uma junção entre essas duas tabelas. A junção é calculada no cluster e retornada. A junção não ocorre no DynamoDB. Este exemplo retorna uma lista de clientes e suas compras para clientes que efetuaram mais de dois pedidos.

```

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;

```

## Para unir duas tabelas de diferentes origens

- No exemplo a seguir, *Customer\_S3* é uma tabela do Hive que carrega um arquivo armazenado CSV no Amazon S3 e *hive\_purchases* é uma tabela que faz referência a dados no DynamoDB.

O exemplo a seguir une dados de clientes armazenados como um CSV arquivo no Amazon S3 com dados de pedidos armazenados no DynamoDB para retornar um conjunto de dados que representa pedidos feitos por clientes que têm “Miller” em seu nome.

```
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

Nos exemplos anteriores, as CREATE TABLE declarações foram incluídas em cada exemplo para maior clareza e integridade. Durante a execução de várias consultas ou operações de exportação em uma determinada tabela do Hive, você só precisa criar a tabela uma vez, no início da sessão do Hive.

## Otimizando o desempenho das EMR operações da Amazon no DynamoDB

EMRAs operações da Amazon em uma tabela do DynamoDB contam como operações de leitura e estão sujeitas às configurações de taxa de transferência provisionada da tabela. A Amazon EMR implementa sua própria lógica para tentar equilibrar a carga em sua tabela do DynamoDB para minimizar a possibilidade de exceder sua taxa de transferência provisionada. Ao final de cada consulta do Hive, a Amazon EMR retorna informações sobre o cluster usado para processar

a consulta, incluindo quantas vezes sua taxa de transferência provisionada foi excedida. Você pode usar essas informações, bem como CloudWatch métricas sobre a taxa de transferência do DynamoDB, para gerenciar melhor a carga na tabela do DynamoDB em solicitações subsequentes.

Os seguintes fatores influenciam a performance de consultas do Hive durante o trabalho com tabelas do DynamoDB.

## Unidades de capacidade de leitura provisionadas

Quando você executa consultas do Hive em uma tabela do DynamoDB, precisa certificar-se de ter provisionado uma quantidade suficiente de unidades de capacidade de leitura.

Por exemplo, suponha que você tenha provisionado 100 unidades de capacidade de leitura para a sua tabela do DynamoDB. Isso permitirá que você realize 100 leituras, ou 409.600 bytes, por segundo. Se essa tabela contiver 20 GB de dados (21.474.836.480 bytes) e sua consulta do Hive realizar uma verificação de tabela completa, você poderá estimar quanto tempo a consulta demorará para ser executada:

$$21.474.836.480/409.600 = 52.429 \text{ segundos} = 14,56 \text{ horas}$$

A única maneira de diminuir o tempo necessário seria ajustar as unidades de capacidade de leitura na tabela DynamoDB de origem. Adicionar mais nós ao EMR cluster da Amazon não ajudará.

Na saída do Hive, a porcentagem de conclusão é atualizada quando um ou mais processos de mapeador são finalizados. Para uma tabela grande do DynamoDB com uma configuração baixa de Capacidade de leitura provisionada, a saída de percentual de conclusão pode não ser atualizada por um longo tempo. No caso acima, o trabalho parecerá estar 0% concluído por várias horas. Para obter um status mais detalhado sobre o progresso do seu trabalho, acesse o EMR console da Amazon; você poderá visualizar o status individual da tarefa do mapeador e as estatísticas das leituras de dados.

Você também pode fazer logon na interface do Hadoop no nó principal e visualizar as estatísticas do Hadoop. Ela mostra o status de tarefas de mapa individuais e algumas estatísticas de leitura de dados. Para obter mais informações, consulte [Interfaces da Web hospedadas no nó principal](#) no Amazon EMR Management Guide.

## Configuração de porcentagem de leitura

Por padrão, EMR a Amazon gerencia a carga de solicitações em relação à sua tabela do DynamoDB de acordo com sua taxa de transferência provisionada atual. No entanto, quando a Amazon EMR retorna informações sobre seu trabalho que incluem um alto número de respostas excedidas



na taxa de transferência provisionada, você pode ajustar a taxa de leitura padrão usando o `dynamodb.throughput.read.percent` parâmetro ao configurar a tabela do Hive. Para obter mais informações sobre como configurar o parâmetro de percentual de leitura, consulte [Opções do Hive](#).

## Configuração de porcentagem de gravação

Por padrão, EMR a Amazon gerencia a carga de solicitações em relação à sua tabela do DynamoDB de acordo com sua taxa de transferência provisionada atual. No entanto, quando a Amazon EMR retorna informações sobre seu trabalho que incluem um alto número de respostas excedidas na taxa de transferência provisionada, você pode ajustar a taxa de gravação padrão usando o `dynamodb.throughput.write.percent` parâmetro ao configurar a tabela do Hive. Para obter mais informações sobre como configurar o parâmetro de percentual de gravação, consulte [Opções do Hive](#).

## Configuração de duração da repetição

Por padrão, a Amazon EMR executa novamente uma consulta do Hive se ela não retornar um resultado em dois minutos, o intervalo padrão de repetição. É possível ajustar esse intervalo definindo o parâmetro `dynamodb.retry.duration` ao executar uma consulta do Hive. Para obter mais informações sobre como configurar o parâmetro de percentual de gravação, consulte [Opções do Hive](#).

## Número de tarefas map

Os daemons mapeadores que o Hadoop executa para processar suas solicitações de exportação e consulta de dados armazenados no DynamoDB estão limitado a uma taxa de leitura máxima de 1 MiB por segundo, para limitar a capacidade de leitura utilizada. Se você tiver throughput provisionado adicional disponível no DynamoDB, poderá melhorar a performance de operações de exportação e consulta do Hive aumentando o número de daemons mapeadores. Para fazer isso, você pode aumentar o número de EC2 instâncias em seu cluster ou aumentar o número de daemons do mapeador em execução em cada instância. EC2

Você pode aumentar o número de EC2 instâncias em um cluster interrompendo o cluster atual e reiniciando-o com um número maior de EC2 instâncias. Você especifica o número de EC2 instâncias na caixa de diálogo Configurar EC2 instâncias se estiver iniciando o cluster a partir do EMR console da Amazon ou com a `--num-instances` opção se estiver iniciando o cluster a partir do CLI.

O número de tarefas de mapeamento executadas em uma instância depende do tipo de EC2 instância. Para obter mais informações sobre os tipos de EC2 instância compatíveis e o número de

mapeadores que cada um fornece, consulte [Configuração da tarefa](#). Lá, você encontrará uma seção "Configuração de tarefas" para cada uma das configurações com suporte.

Outra maneira de aumentar o número de daemons mapeadores é alterar o parâmetro de configuração `mapreduce.tasktracker.map.tasks.maximum` do Hadoop para um valor mais alto. Isso tem a vantagem de oferecer mais mapeadores sem aumentar o número ou o tamanho das EC2 instâncias, o que economiza dinheiro. Uma desvantagem é que definir esse valor muito alto pode fazer com que as EC2 instâncias do seu cluster fiquem sem memória. Para definir `mapreduce.tasktracker.map.tasks.maximum`, inicie o cluster e especifique um valor para `mapreduce.tasktracker.map.tasks.maximum` como uma propriedade da classificação de configuração `mapred-site`. Isso é mostrado no exemplo a seguir. Para obter mais informações, consulte [Configurar aplicações](#).

```
{
 "configurations": [
 {
 "classification": "mapred-site",
 "properties": {
 "mapred.tasktracker.map.tasks.maximum": "10"
 }
 }
]
}
```

## Solicitações de dados em paralelo

Várias solicitações de dados, seja de mais de um usuário ou de mais de um aplicativo, para uma única tabela podem esgotar o throughput provisionado de leitura e diminuir o desempenho.

## Duração do processo

A consistência dos dados no DynamoDB depende da ordem de operações de leitura e gravação em cada nó. Embora uma consulta do Hive esteja em andamento, outra aplicação pode carregar novos dados para a tabela do DynamoDB ou modificar ou excluir dados existentes. Nesse caso, os resultados da consulta do Hive podem não refletir as alterações feitas nos dados enquanto a consulta estava em execução.

## Evitar exceder o throughput

Ao executar consultas do Hive no DynamoDB, tome cuidado para não exceder seu throughput provisionado, pois isso esgotará a capacidade necessária para chamadas da aplicação para

DynamoDB : :Get. Para garantir que isso não ocorra, você deve monitorar regularmente o volume de leitura e a limitação das chamadas do aplicativo, DynamoDB : :Get verificando os registros e as métricas de monitoramento na Amazon. CloudWatch

## Tempo de solicitação

Programas consultas do Hive que acessam uma tabela do Dynamo quando há menor demanda na tabela do DynamoDB melhora a performance. Por exemplo, se a maioria dos usuários do aplicativo morar em São Francisco, você poderá optar por exportar os dados diariamente às 4h. PST, quando a maioria dos usuários está dormindo e não atualizando registros no banco de dados do DynamoDB.

## Tabelas baseadas em tempo

Se os dados estiverem organizados como uma série de tabelas do DynamoDB com base no tempo, como uma tabela por dia, você poderá exportá-los quando a tabela não estiver mais ativa. Você pode usar essa técnica para fazer o backup dos dados para o Amazon S3 de uma forma contínua.

## Dados arquivados

Se você planeja executar muitas consultas do Hive nos dados armazenados no DynamoDB e seu aplicativo pode tolerar dados arquivados, convém exportar os dados para o Amazon HDFS S3 e executar as consultas do Hive em uma cópia dos dados em vez do DynamoDB. Isto conserva suas operações de leitura e o throughput provisionado.

## Kinesis

EMROs clusters da Amazon podem ler e processar streams do Amazon Kinesis diretamente, usando ferramentas conhecidas no ecossistema Hadoop, como Hive, Pig MapReduce, Hadoop Streaming e Cascading. API Você também pode unir dados em tempo real do Amazon Kinesis com dados existentes no Amazon S3, no Amazon DynamoDB e em um cluster em execução. HDFS Você pode carregar diretamente os dados da Amazon EMR para o Amazon S3 ou o DynamoDB para atividades de pós-processamento. Para obter informações sobre os destaques e o preço do serviço Amazon Kinesis, consulte a página do [Amazon Kinesis](#).

## O que posso fazer com a integração entre a Amazon EMR e o Amazon Kinesis?

A integração entre o Amazon EMR e o Amazon Kinesis facilita muito certos cenários; por exemplo:

- **Análise de log de transmissão:** você pode analisar logs Web de transmissão para gerar uma lista dos dez maiores tipos de erros, em intervalos de minutos, por região, navegador e domínio de acesso.
- **Envolvimento do cliente:** você pode criar consultas que unam os dados de clickstream do Amazon Kinesis com as informações de campanhas publicitárias armazenadas em uma tabela do DynamoDB para identificar as mais eficientes categorias de anúncios que são exibidos em determinados sites.
- **Consultas interativas ad-hoc** — Você pode carregar periodicamente dados dos streams do Amazon Kinesis e disponibilizá-los como uma tabela local do Impala para consultas analíticas rápidas HDFS e interativas.

## Análise com ponto de verificação de fluxos do Amazon Kinesis

Os usuários podem executar análises periódicas em lote de fluxos do Amazon Kinesis no que chamados de iterações. Como os registros de dados de stream do Amazon Kinesis são recuperados usando um número de sequência, os limites de iteração são definidos pelos números de sequência inicial e final que a Amazon armazena EMR em uma tabela do DynamoDB. Por exemplo, quando `iteration0` é encerrado, ele armazena o número de sequência final no DynamoDB, para que, quando o trabalho `iteration1` começar, ele possa recuperar dados subsequentes do fluxo. Esse mapeamento de iterações em dados de stream é chamado de pontos de verificação. Para obter mais informações, consulte [Kinesis connector](#).

Se uma iteração foi verificada e o trabalho falhou no processamento de uma iteração, a Amazon EMR tentará reprocessar os registros nessa iteração.

Pontos de verificação são um recurso que permite:

- Iniciar o processamento de dados após um número de sequência processado por uma consulta anterior que foi executada no mesmo stream e nome lógico
- Reprocessar o mesmo lote de dados do Kinesis que foi processado por uma consulta anterior

Para habilitar pontos de verificação, defina o parâmetro `kinesis.checkpoint.enabled` como `true` nos seus scripts. Além disso, configure os seguintes parâmetros:

| Definição da configuração                                 | Descrição                                                                                                                                  |
|-----------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------|
| <code>kinesis.checkpoint.metastore.table.name</code>      | Nome da tabela do DynamoDB onde serão armazenadas informações de ponto de verificação                                                      |
| <code>kinesis.checkpoint.metastore.hash.key.name</code>   | Nome da chave de hash da tabela do DynamoDB                                                                                                |
| <code>kinesis.checkpoint.metastore.hash.range.name</code> | Nome da chave de intervalo da tabela do DynamoDB                                                                                           |
| <code>kinesis.checkpoint.logical.name</code>              | Um nome lógico para o processamento atual                                                                                                  |
| <code>kinesis.checkpoint.iteration.no</code>              | O número de iterações para o processamento associado ao nome lógico                                                                        |
| <code>kinesis.rerun.iteration.without.wait</code>         | Valor booleano que indica se uma iteração com falha pode ser executada novamente sem esperar o tempo limite; o padrão é <code>false</code> |

## IOPS Recomendações provisionadas para tabelas do Amazon DynamoDB

O EMR conector Amazon para Amazon Kinesis usa o banco de dados do DynamoDB como suporte para metadados de pontos de verificação. Você deve criar uma tabela no DynamoDB antes de consumir dados em um stream do Amazon Kinesis com um EMR cluster da Amazon em intervalos de verificação. A tabela deve estar na mesma região do seu EMR cluster da Amazon. A seguir estão recomendações gerais sobre o número de pessoas IOPS que você deve provisionar para suas tabelas do DynamoDB;  $j$  seja o número máximo de trabalhos do Hadoop (com uma combinação diferente de nome lógico e número de iteração) que podem ser executados simultaneamente e  $s$  seja o número máximo de fragmentos que qualquer trabalho processará:

Para Read Capacity Units (Unidades de capacidade de leitura):  $j*s/5$

Para Write Capacity Units (Unidades de capacidade de gravação):  $j*s$

## Considerações sobre a performance

A taxa de transferência do fragmento do Amazon Kinesis é diretamente proporcional ao tamanho da instância dos nós nos EMR clusters da Amazon e ao tamanho do registro no stream. Recomendamos que você use m5.xlarge ou instâncias maiores no nó principal e nos nós centrais.

## Agende análises do Amazon Kinesis com a Amazon EMR

Quando você estiver analisando dados em um fluxo do Amazon Kinesis ativo, limitado por tempos limite e uma duração máxima para qualquer iteração, será importante executar sempre a análise para coletar detalhes periódicos do fluxo. Existem várias maneiras de executar esses scripts e consultas em intervalos periódicos; recomendamos usar AWS Data Pipeline em tarefas recorrentes como essas. Para obter mais informações, consulte [AWS Data Pipeline PigActivity](#) e [AWS Data Pipeline HiveActivity](#) no Guia do AWS Data Pipeline desenvolvedor.

## Migração do conector Spark Kinesis para SDK 2.x para Amazon 7.0 EMR

O AWS SDK fornece um rico conjunto de APIs bibliotecas para interagir com serviços de computação em AWS nuvem, como gerenciamento de credenciais, conexão com serviços S3 e Kinesis. O conector do Spark Kinesis é usado para consumir dados do Kinesis Data Streams, e os dados recebidos são transformados e processados no mecanismo de execução do Spark. Atualmente, esse conector é construído sobre 1.x de AWS SDK e Kinesis-client-library (KCL).

Como parte da migração AWS SDK 2.x, o conector Spark Kinesis também é atualizado adequadamente para ser executado com o 2.x. SDK Na versão do Amazon EMR 7.0, o Spark contém a atualização SDK 2.x que ainda não está disponível na versão comunitária do Apache Spark. Se você usa o conector Spark Kinesis de uma versão inferior à 7.0, você deve migrar os códigos do seu aplicativo para serem executados SDK na versão 2.x antes de poder migrar para a Amazon 7.0. EMR

## Guias de migração

Esta seção descreve as etapas para migrar uma aplicação ao conector atualizado do Spark Kinesis. Ele inclui guias para migrar para a Kinesis Client Library KCL () 2.x AWS , provedores de credenciais AWS e clientes de serviços na versão 2.x. AWS SDK Para referência, ele também inclui um [WordCount](#) programa de amostra que usa o conector Kinesis.

### Tópicos

- [Migrando KCL de 1.x para 2.x](#)

- [Migração de provedores de AWS credenciais de AWS SDK 1.x para 2.x](#)
- [Migração de clientes AWS de serviço de AWS SDK 1.x para 2.x](#)
- [Exemplos de códigos para aplicações de streaming](#)
- [Considerações ao usar o conector atualizado do Spark Kinesis](#)

## Migrando KCL de 1.x para 2.x

- Nível e dimensões das métricas em **KinesisInputDStream**

Ao instanciar um `KinesisInputDStream`, você pode controlar o nível e as dimensões das métricas do fluxo. O exemplo a seguir demonstra como você pode personalizar esses parâmetros com 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()
```

Na KCL versão 2.x, essas configurações têm nomes de pacotes diferentes. Para migrar à versão 2.x:

1. Altere as instruções de importação de `com.amazonaws.services.kinesis.clientlibrary.lib.worker.KinesisClientLibCon` e `com.amazonaws.services.kinesis.metrics.interfaces.MetricsLevel` para `software.amazon.kinesis.metrics.MetricsLevel` e `software.amazon.kinesis.metrics.MetricsUtil`, respectivamente.

```
// 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. Substitua a linha

```
metricsEnabledDimensionsKinesisClientLibConfiguration.DEFAULT_METRICS_ENABLED_DIMENSIONS
por metricsEnabledDimensionsSet(MetricsUtil.OPERATION_DIMENSION_NAME,
MetricsUtil.SHARD_ID_DIMENSION_NAME)
```

A seguir está uma versão atualizada de `KinesisInputDStream` com níveis e dimensões de métricas personalizados.

```
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()
```

- Função de manipulador de mensagens em `KinesisInputDStream`

Ao instanciar um `KinesisInputDStream`, você também pode fornecer uma “função de manipulador de mensagens” que usa um registro do Kinesis e retorna um objeto genérico `T`, caso queira usar outros dados incluídos em um registro, como a chave de partição.

Na KCL versão 1.x, a assinatura da função do manipulador de mensagens é: `Record => T`, onde `Record` está. `com.amazonaws.services.kinesis.model.Record` Na KCL versão



2.x, a assinatura do manipulador é alterada para: `KinesisClientRecord => T`, where `KinesisClientRecord` is `software.amazon.kinesis.retrieval.KinesisClientRecord`

A seguir está um exemplo de fornecimento de um manipulador de mensagens na KCL versão 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)
```

Para migrar o manipulador de mensagens:

1. Altere a instrução de importação de `com.amazonaws.services.kinesis.model.Record` para `software.amazon.kinesis.retrieval.KinesisClientRecord`.

```
// import com.amazonaws.services.kinesis.model.Record
import software.amazon.kinesis.retrieval.KinesisClientRecord
```

2. Atualize a assinatura do método do manipulador de mensagens.

```
//def addFive(r: Record): Int = JavaUtils.bytesToString(r.getData).toInt + 5
def addFive = (r: KinesisClientRecord) => JavaUtils.bytesToString(r.data()).toInt
+ 5
```

A seguir está um exemplo atualizado de fornecimento do manipulador de mensagens na KCL versão 2.x:

```
import software.amazon.kinesis.retrieval.KinesisClientRecord
```

```
def addFive = (r: KinesisClientRecord) => JavaUtils.byteToString(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)
```

Para obter mais informações sobre a migração de KCL 1.x para 2.x, consulte [Migração de consumidores de KCL 1.x para 2.x](#). KCL

## Migração de provedores de AWS credenciais de AWS SDK 1.x para 2.x

Os provedores de credenciais são usados para obter AWS credenciais para interações com. [AWS](#) [Há várias mudanças de interface e classe relacionadas aos provedores de credenciais na SDK versão 2.x, que podem ser encontradas aqui](#). O conector do Spark Kinesis definiu uma interface `org.apache.spark.streaming.kinesis.SparkAWSCredentials` () e classes de implementação que retornam a versão AWS 1.x dos provedores de credenciais. Esses provedores de credenciais são necessários ao inicializar clientes Kinesis. Por exemplo, se você estiver usando o método `SparkAWSCredentials.provider` nos aplicativos, precisará atualizar os códigos para consumir a versão 2.x dos provedores de AWS credenciais.

Veja a seguir um exemplo do uso dos provedores de credenciais na AWS SDK versão 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")
```

Para migrar para a SDK versão 2.x:

1. Altere a instrução de importação de `com.amazonaws.auth.AWSCredentialsProvider` para `software.amazon.awssdk.auth.credentials.AwsCredentialsProvider`.

```
//import com.amazonaws.auth.AWSCredentialsProvider
import software.amazon.awssdk.auth.credentials.AwsCredentialsProvider
```

2. Atualize os códigos restantes que usam essa 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")
```

Migração de clientes AWS de serviço de AWS SDK 1.x para 2.x

AWS clientes de serviço têm nomes de pacotes diferentes em 2.x (ou seja, `software.amazon.awssdk`), enquanto o SDK 1.x usa `com.amazonaws`. Para obter mais informações sobre as alterações de clientes, consulte [aqui](#). Se você estiver usando esses clientes de serviços nos códigos, precisará migrá-los adequadamente.

Veja a seguir um exemplo de criação de um cliente na SDK versão 1.x:

```
import com.amazonaws.services.dynamodbv2.AmazonDynamoDBClient
import com.amazonaws.services.dynamodbv2.document.DynamoDB

AmazonDynamoDB ddbClient = AmazonDynamoDBClientBuilder.defaultClient();
AmazonDynamoDBClient ddbClient = new AmazonDynamoDBClient();
```

Para migrar à versão 2.x:

1. Altere as instruções de importação dos clientes de serviços. Veja os clientes `DynamoDB` como exemplo. Você precisaria mudar

`com.amazonaws.services.dynamodbv2.AmazonDynamoDBClient` ou  
`com.amazonaws.services.dynamodbv2.document.DynamoDB` para  
`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. Atualização dos códigos que inicializam os clientes

```
// AmazonDynamoDB ddbClient = AmazonDynamoDBClientBuilder.defaultClient();
// AmazonDynamoDBClient ddbClient = new AmazonDynamoDBClient();

DynamoDbClient ddbClient = DynamoDbClient.create();
DynamoDbClient ddbClient = DynamoDbClient.builder().build();
```

Para obter mais informações sobre a migração AWS SDK de 1.x para 2.x, consulte [Qual é a diferença entre o AWS SDK para Java 1.x e 2.x](#)

## Exemplos de códigos para aplicações 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()
}
}
```

## Considerações ao usar o conector atualizado do Spark Kinesis

- Se seus aplicativos usam a JDK versão `Kinesis-producer-library` com versão inferior a 11, você pode se deparar com exceções como `java.lang.NoClassDefFoundError: javax/xml/bind/DatatypeConverter`. Isso acontece porque o EMR 7.0 vem com JDK 17 por padrão e os módulos J2EE foram removidos das bibliotecas padrão desde o Java 11+. Isso pode ser corrigido adicionando a dependência a seguir no arquivo pom. Substitua a versão da biblioteca pela dependência que preferir.

```
<dependency>
 <groupId>javax.xml.bind</groupId>
 <artifactId>jaxb-api</artifactId>
 <version>${jaxb-api.version}</version>
</dependency>
```

- O jar do conector do Spark Kinesis pode ser encontrado nesse caminho após a criação de EMR um cluster: `/usr/lib/spark/connector/lib/`

## S3 DistCp (s3-dist-cp)

O Apache DistCp é uma ferramenta de código aberto que você pode usar para copiar grandes quantidades de dados. O S3 DistCp é semelhante DistCp, mas otimizado para trabalhar AWS, especialmente com o Amazon S3. O comando para S3 DistCp na Amazon EMR versão 4.0 e posterior é `s3-dist-cp`, que você adiciona como uma etapa em um cluster ou na linha de comando. Usando o S3DistCp, você pode copiar com eficiência grandes quantidades de dados do Amazon S3 HDFS para onde eles podem ser processados por etapas subsequentes em seu cluster da AmazonEMR. Você também pode usar o S3 DistCp para copiar dados entre buckets do Amazon S3 ou do HDFS Amazon S3. O S3 DistCp é mais escalável e eficiente para copiar paralelamente grandes números de objetos entre buckets e contas. AWS

Para comandos específicos que demonstram a flexibilidade do S3DistCP em cenários do mundo real, consulte [Sete dicas para usar o S3 DistCp](#) no blog de Big Data. AWS

Por exemplo DistCp, o S3 DistCp usa MapReduce para copiar de forma distribuída. Ele compartilha a cópia, o tratamento de erros, a recuperação e as tarefas de relatórios com vários servidores. Para obter mais informações sobre o projeto de código DistCp aberto Apache, consulte o [DistCp guia na documentação](#) do Apache Hadoop.

Se o S3 DistCp não conseguir copiar alguns ou todos os arquivos especificados, a etapa do cluster falhará e retornará um código de erro diferente de zero. Se isso ocorrer, o S3 DistCp não limpará os arquivos parcialmente copiados.

### Important

O S3 DistCp não é compatível com nomes de bucket do Amazon S3 que contenham o caractere de sublinhado.

O S3 DistCp não oferece suporte à concatenação de arquivos Parquet. Use PySpark em vez disso. Para obter mais informações, consulte [Concatenação de arquivos de parquet na Amazon](#). EMR

Para evitar erros de cópia ao usar o S3DistCp para copiar um único arquivo (em vez de um diretório) do S3 para, HDFS use a Amazon EMR versão 5.33.0 ou posterior, ou a Amazon versão 6.3.0 ou posterior. EMR



## Opções do S3 DistCp

Embora semelhante DistCp, o S3 DistCp oferece suporte a um conjunto diferente de opções para alterar a forma como ele copia e compacta dados.

Ao chamar o S3DistCp, você pode especificar as opções descritas na tabela a seguir. As opções são adicionadas à etapa, usando-se a lista de argumentos. Exemplos dos DistCp argumentos do S3 são mostrados na tabela a seguir.

Opção	Descrição	Obrigatório
<code>--src=LOCATION</code>	Local dos dados a serem copiados. Isso pode ser um local HDFS ou o Amazon S3.  Exemplo: <code>--src=s3://amzn-s3-demo-bucket1/logs/j-3GYXXXXXX9I0J/node</code>	Sim



Opção	Descrição	Obrigatório
	<p> <b>Important</b></p> <p>O S3 DistCp não é compatível com nomes de bucket do Amazon S3 que contenham o caractere de sublinhado.</p>	
<p><code>--dest=LOCATION</code></p>	<p>Destino para os dados. Isso pode ser um local HDFS ou o Amazon S3.</p> <p>Exemplo: <code>--dest=hdfs:///output</code></p> <p> <b>Important</b></p> <p>O S3 DistCp não é compatível com nomes de bucket do Amazon S3 que contenham o caractere de sublinhado.</p>	<p>Sim</p>
<p><code>--srcPattern=PATTERN</code></p>	<p>Uma <a href="#">expressão regular</a> que filtra a operação de cópia para um subconjunto dos dados em <code>--src</code>. Se nenhum <code>--srcPattern</code> nem <code>--groupBy</code> for especificado, todos os dados em <code>--src</code> serão copiados em <code>--dest</code>.</p> <p>Se o argumento da expressão regular contiver caracteres especiais, como um asterisco (*), a expressão regular ou a string <code>--args</code> inteira deverá ser colocada entre aspas simples (').</p> <p>Exemplo: <code>--srcPattern=.*daemons.*-ha doop-.*</code></p>	<p>Não</p>

Opção	Descrição	Obrigatório
<pre>--groupBy=PATTERN</pre>	<p>Uma <a href="#">expressão regular</a> que faz com que o S3 concatene arquivos que correspondam DistCp à expressão. Por exemplo, você pode usar essa opção para combinar todos os arquivos de log gravados em uma hora em um único arquivo. O nome do arquivo concatenado é o valor correspondido pela expressão regular para o agrupamento.</p> <p>Parênteses indicam como os arquivos devem ser agrupados, com todos os itens que correspondam à instrução parentética sendo combinados em um único arquivo de saída. Se a expressão regular não incluir uma declaração entre parênteses, o cluster falhará na DistCp etapa S3 e retornará um erro.</p> <p>Se o argumento da expressão regular contiver caracteres especiais, como um asterisco (*), a expressão regular ou a string <code>--args</code> inteira deverá ser colocada entre aspas simples (').</p> <p>Quando <code>--groupBy</code> for especificado, somente os arquivos que correspondam ao padrão especificado serão copiados. Não é necessário especificar <code>--groupBy</code> e <code>--srcPattern</code> ao mesmo tempo.</p> <p>Exemplo: <code>--groupBy=.*subnetid.*([0-9]+-[0-9]+-[0-9]+-[0-9]+)</code>.</p>	Não

Opção	Descrição	Obrigatório
<code>--targetSize=SIZE</code>	<p>O tamanho, em mebibytes (MiB), dos arquivos a serem criados com base na opção <code>--groupBy</code>. Esse valor deve ser um número inteiro. Quando <code>--targetSize</code> definido, o S3 DistCp tenta corresponder a esse tamanho; o tamanho real dos arquivos copiados pode ser maior ou menor que esse valor. Os trabalhos são agregados com base no tamanho do arquivo de dados, portanto, é possível que o tamanho do arquivo de destino corresponda ao tamanho do arquivo de dados de origem.</p> <p>Se os arquivos concatenados pelo <code>--groupBy</code> forem maiores que o valor do <code>--targetSize</code>, eles serão divididos em arquivos de parte e nomeados sequencialmente com um valor numérico anexado ao final. Por exemplo, um arquivo concatenado em <code>myfile.gz</code> seria dividido em partes como: <code>myfile0.gz</code>, <code>myfile1.gz</code> etc.</p> <p>Exemplo: <code>--targetSize=2</code></p>	Não
<code>--appendToLastFile</code>	<p>Especifica o comportamento do S3 DistCp ao copiar para arquivos do Amazon S3 que já estão HDFS presentes. Ele acrescenta novos dados de arquivos aos arquivos existentes. Se você usar <code>--appendToLastFile</code> com <code>--groupBy</code>, novos dados serão anexados aos arquivos que correspondam aos mesmos grupos. Essa opção também respeita o comportamento <code>--targetSize</code> quando usada com <code>--groupBy</code>.</p>	Não

Opção	Descrição	Obrigatório
<code>--outputCodec=CODEC</code>	<p>Especifica o codec de compactação a ser usado para os arquivos copiados. Isso pode ter os valores: <code>gzip</code>, <code>gz</code>, <code>lzo</code>, <code>snappy</code> ou <code>none</code>. Você pode usar essa opção, por exemplo, para converter arquivos de entrada compactados com Gzip em arquivos de saída com LZO compactação ou para descompactar os arquivos como parte da operação de cópia. Se você selecionar um codec de saída, o nome do arquivo será anexado com a extensão apropriada (por exemplo, para <code>gz</code> e <code>gzip</code>, a extensão é <code>.gz</code>) Se você não especificar um valor para <code>--outputCodec</code>, os arquivos serão copiados sem alterações na compactação.</p> <p>Exemplo: <code>--outputCodec=lzo</code></p>	Não
<code>--s3ServerSideEncryption</code>	<p>Garante que os dados de destino sejam transferidos usando SSL e criptografados automaticamente no Amazon S3 usando uma chave do lado do AWS serviço. Ao recuperar dados usando o S3DistCp, os objetos são automaticamente descriptografados. Se você tentar copiar um objeto não criptografado em um bucket do Amazon S3 que exige criptografia, haverá falha na operação. Para obter mais informações, consulte <a href="#">Proteger dados com criptografia</a>.</p> <p>Exemplo: <code>--s3ServerSideEncryption</code></p>	Não

Opção	Descrição	Obrigatório
<code>--deleteOnSuccess</code>	<p>Se a operação de cópia for bem-sucedida, essa opção fará com que o S3 DistCp exclua os arquivos copiados do local de origem. Isso é útil se você estiver copiando arquivos de saída, como arquivos de log, de um local para outro como uma tarefa programada e não quiser copiar os mesmos arquivos duas vezes.</p> <p>Exemplo: <code>--deleteOnSuccess</code></p>	Não
<code>--disableMultipartUpload</code>	<p>Desativa o uso do multipart upload.</p> <p>Exemplo: <code>--disableMultipartUpload</code></p>	Não
<code>--multipartUploadChunkSize=SIZE</code>	<p>O tamanho, em MiB, de cada parte de um carregamento multiparte do Amazon S3. O S3 DistCp usa upload de várias partes quando copia dados maiores que o <code>multipartUploadChunkSize</code>. Para melhorar a performance do trabalho, você pode aumentar o tamanho de cada parte. O tamanho padrão é 128 MiB.</p> <p>Exemplo: <code>--multipartUploadChunkSize=1000</code></p>	Não
<code>--numberOfFiles</code>	<p>Precede os arquivos de saída com números sequenciais. A contagem inicia em 0, a menos que um valor diferente seja especificado por <code>--startingIndex</code>.</p> <p>Exemplo: <code>--numberOfFiles</code></p>	Não
<code>--startingIndex=INDEX</code>	<p>Usado com <code>--numberOfFiles</code> para especificar o primeiro número na sequência.</p> <p>Exemplo: <code>--startingIndex=1</code></p>	Não

Opção	Descrição	Obrigatório
<pre>--outputManifest=FILENAME</pre>	<p>Cria um arquivo de texto, compactado com Gzip, que contém uma lista de todos os arquivos copiados pelo S3. DistCp</p> <p>Exemplo: <code>--outputManifest=manifest-1.gz</code></p>	Não
<pre>--previousManifest=PATH</pre>	<p>Lê um arquivo de manifesto que foi criado durante uma chamada anterior para o S3 DistCp usando o <code>--outputManifest</code> sinalizador. Quando o <code>--previousManifest</code> sinalizador é definido, o S3 DistCp exclui os arquivos listados no manifesto da operação de cópia. Se <code>--outputManifest</code> for especificado juntamente com <code>--previousManifest</code>, os arquivos listados no manifesto anterior também aparecerão no novo arquivo manifesto, embora os arquivos não sejam copiados.</p> <p>Exemplo: <code>--previousManifest=/usr/bin/manifest-1.gz</code></p>	Não
<pre>--requirePreviousManifest</pre>	<p>Requer um manifesto anterior criado durante uma chamada anterior para o S3DistCp. Se isso for definido como falso, nenhum erro será gerado quando um manifesto anterior não for especificado. O padrão é true.</p>	Não
<pre>--copyFromManifest</pre>	<p>Inverte o comportamento de <code>--previousManifest</code> fazer com que o S3 DistCp use o arquivo de manifesto especificado como uma lista de arquivos a serem copiados, em vez de uma lista de arquivos a serem excluídos da cópia.</p> <p>Exemplo: <code>--copyFromManifest --previousManifest=/usr/bin/manifest-1.gz</code></p>	Não

Opção	Descrição	Obrigatório
<code>--s3Endpoint=ENDPOINT</code>	<p>Especifica o endpoint do Amazon S3 a ser usado ao carregar um arquivo. Essa opção define o endpoint para a origem e o destino. Se não estiver definido, o endpoint padrão será <code>s3.amazonaws.com</code>. Para obter uma lista dos endpoints do Amazon S3, consulte <a href="#">Regions and endpoints</a>.</p> <p>Exemplo: <code>--s3Endpoint=s3.eu-west-1.amazonaws.com</code></p>	Não
<code>--storageClass=CLASS</code>	<p>A classe de armazenamento a ser usada quando o destino é o Amazon S3. Os valores válidos são <code>STANDARD</code> e <code>REDUCED_REDUNDANCY</code>. Se essa opção não for especificada, o S3 DistCp tentará preservar a classe de armazenamento.</p> <p>Exemplo: <code>--storageClass=STANDARD</code></p>	Não

Opção	Descrição	Obrigatório
<pre>--srcPrefixesFile=PATH</pre>	<p>um arquivo de texto no Amazon S3 (s3://), HDFS (hdfs://) ou sistema de arquivos local (file://) que contém uma lista de <code>src</code> prefixos, um prefixo por linha.</p> <p>Se <code>srcPrefixesFile</code> for fornecido, o S3 não DistCp listará o caminho <code>src</code>. Em vez disso, ele gerará uma lista de fontes como resultado combinado da listagem de todos os prefixos especificados nesse arquivo. O caminho relativo em comparação com o caminho <code>src</code>, em vez desses prefixos, será usado para gerar os caminhos de destino. Se <code>srcPattern</code> também for especificado, ele será aplicado à lista de resultados combinados dos prefixos de origem para filtrar ainda mais a entrada. Se <code>copyFromManifest</code> for usado, os objetos no manifesto serão copiados e <code>srcPrefixesFile</code> será ignorado.</p> <p>Exemplo: <code>--srcPrefixesFile=PATH</code></p>	Não

Além das opções acima, o S3 DistCp implementa a [interface da ferramenta](#), o que significa que ela suporta as opções genéricas.

## Adicionando o S3 DistCp como uma etapa em um cluster

Você pode chamar o S3 DistCp adicionando-o como uma etapa em seu cluster. As etapas podem ser adicionadas a um cluster na inicialização ou a um cluster em execução usando o console, CLI, ou API. Os exemplos a seguir demonstram a adição de uma DistCp etapa do S3 a um cluster em execução. Para obter mais informações sobre como adicionar etapas a um cluster, consulte [Enviar trabalho para um cluster](#) no Amazon EMR Management Guide.



Para adicionar uma DistCp etapa do S3 a um cluster em execução usando o AWS CLI

Para obter mais informações sobre o uso dos EMR comandos da Amazon no AWS CLI, consulte a [Referência de AWS CLI comandos](#).

- Para adicionar uma etapa a um cluster que chama o S3DistCp, passe os parâmetros que especificam como o S3 DistCp deve realizar a operação de cópia como argumentos.

O exemplo a seguir copia logs do daemon do Amazon S3 para `hdfs:///output`. No comando a seguir:

- `--cluster-id` especifica o cluster
- `Jar` é a localização do DistCp JAR arquivo S3. Para ver um exemplo de como executar um comando em um cluster usando o `command-runner.jar`, consulte [Enviar uma JAR etapa personalizada para executar um script ou comando](#).
- `Args` é uma lista separada por vírgulas dos pares nome-valor da opção a serem passados para o S3. DistCp Para obter uma lista completa das opções disponíveis, consulte [Opções do S3 DistCp](#).

Para adicionar uma etapa de DistCp cópia do S3 a um cluster em execução, coloque o seguinte em um JSON arquivo salvo no Amazon S3 ou em seu sistema de arquivos local, *myStep.json* como neste exemplo. Substituir *j-3GYXXXXXX9I0K* com seu ID de cluster e substitua *mybucket* com o nome do seu bucket do 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
```

## Example Copiar arquivos de log do Amazon S3 para HDFS

Este exemplo também ilustra como copiar arquivos de log armazenados em um bucket do Amazon S3 adicionando uma etapa a um cluster HDFS em execução. Neste exemplo, a opção `--srcPattern` é usada para limitar os dados copiados para os logs do daemon.

Para copiar arquivos de log do Amazon S3 para HDFS usar a `--srcPattern` opção, coloque o seguinte em um JSON arquivo salvo no Amazon S3 ou em seu sistema de arquivos local, *myStep.json* como neste exemplo. Substituir *j-3GYXXXXXXXX9I0K* com seu ID de cluster e substitua *mybucket* com o nome do seu bucket do Amazon S3.

```
[
 {
 "Name": "S3DistCp step",
 "Args": ["s3-dist-cp", "--s3Endpoint=s3.amazonaws.com", "--src=s3://mybucket/logs/j-3GYXXXXXXXX9I0J/node/", "--dest=hdfs:///output", "--srcPattern=.*daemons.*-hadoop-.*"],
 "ActionOnFailure": "CONTINUE",
 "Type": "CUSTOM_JAR",
 "Jar": "command-runner.jar"
 }
]
```

## Limpendo após falhas em trabalhos do S3 DistCp

Se o S3 DistCp não puder copiar alguns ou todos os arquivos especificados, a etapa de comando ou cluster falhará e retornará um código de erro diferente de zero. Se isso ocorrer, o S3 DistCp não limpará os arquivos parcialmente copiados. Você deverá excluí-los manualmente.

Os arquivos parcialmente copiados são salvos no HDFS tmp diretório em subdiretórios com o identificador exclusivo da tarefa do S3. DistCp Você pode encontrar esse ID na saída padrão do trabalho.

Por exemplo, para um DistCp trabalho do S3 com o ID `4b1c37bb-91af-4391-aaf8-46a6067085a6`, você pode se conectar ao nó principal do cluster e executar o comando a seguir para visualizar os arquivos de saída associados ao trabalho.

```
hdfs dfs -ls /tmp/4b1c37bb-91af-4391-aaf8-46a6067085a6/output
```

O comando retorna uma lista de arquivos semelhantes aos seguintes:

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
```

Você pode executar o comando a seguir para excluir o diretório e todo o conteúdo.

```
hdfs dfs rm -rf /tmp/4b1c37bb-91af-4391-aaf8-46a6067085a6
```

# Executar comandos e scripts em um cluster do Amazon EMR

Este tópico aborda como executar um comando ou script como uma etapa no cluster. Executar um comando ou script como uma etapa é uma das várias maneiras de [enviar trabalhos a um cluster](#) e serve para as seguintes situações:

- Quando você não tem acesso SSH ao cluster do Amazon EMR
- Quando você quer executar um comando bash ou shell para solucionar problemas no cluster

Você pode executar um script ao criar o cluster ou quando o cluster estiver no estado WAITING. Para executar um script antes do início do processamento de etapas, em vez disso, use uma ação de bootstrap. Para obter mais informações sobre ações de bootstrap, consulte [Create bootstrap actions to install additional software](#) no Guia de gerenciamento do Amazon EMR.

O Amazon EMR fornece as seguintes ferramentas que ajudam a executar scripts, comandos e outros programas no cluster. É possível invocar as duas ferramentas usando o console de gerenciamento do Amazon EMR ou o a AWS CLI.

`command-runner.jar`

Localizado na AMI do Amazon EMR do cluster. Você pode usar `command-runner.jar` para executar comandos no cluster. Especifique `command-runner.jar` sem usar o caminho completo.

`script-runner.jar`

Hospedado no Amazon S3, em `s3://<region>.elasticmapreduce/libs/script-runner/script-runner.jar`, onde `<region>` está a região em que o cluster do Amazon EMR reside. Você pode usar `script-runner.jar` para executar scripts salvos no local ou no Amazon S3 em seu cluster. É necessário especificar o URI completo de `script-runner.jar` ao enviar uma etapa.

# Enviar uma etapa JAR personalizada para executar um script ou comando

Os exemplos da AWS CLI a seguir ilustram alguns casos de uso comuns de `command-runner.jar` e `script-runner.jar` no Amazon EMR.

Example : Executar um comando em um cluster usando **command-runner.jar**

Ao usar `command-runner.jar`, você especifica comandos, opções e valores na lista de argumentos da etapa.

O exemplo da AWS CLI a seguir envia uma etapa a um cluster em execução que invoca `command-runner.jar`. O comando especificado na lista `Args` baixa um script chamado *my-script.sh* do Amazon S3 para o diretório inicial do usuário do Hadoop. Em seguida, o comando modifica as permissões do script e executa *my-script.sh*.

Quando você usa a AWS CLI, os itens da lista `Args` devem ser separados por vírgula, sem espaços em branco entre os elementos da lista. Por exemplo, `Args=[example-command,example-option,"example option value"]` em vez de `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"]
```

Example : executar um script em um cluster usando **script-runner.jar**

Ao usar `script-runner.jar`, você especifica o script que deseja executar na lista de argumentos da etapa.

O exemplo da AWS CLI a seguir envia uma etapa a um cluster em execução que invoca `script-runner.jar`. Nesse caso, o script chamado *my-script.sh* é armazenado no Amazon S3. Também é possível especificar scripts locais que são armazenados no nó principal do 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]
```

## Outras formas de usar `command-runner.jar`

Você também pode usar `command-runner.jar` para enviar trabalhos para um cluster usando ferramentas como `spark-submit` ou `hadoop-streaming`. Ao iniciar uma aplicação usando `command-runner.jar`, você especifica `CUSTOM_JAR` como o tipo de etapa em vez de usar um valor como `SPARK`, `STREAMING` ou `PIG`. A disponibilidade da ferramenta varia de acordo com as aplicações instaladas no cluster.

O comando de exemplo a seguir utiliza `command-runner.jar` para enviar uma etapa usando `spark-submit`. A lista `Args` especifica `spark-submit` como comando, seguido pelo URI do Amazon S3 da aplicação Spark `my-app.py` com argumentos e valores.

```
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]
```

A tabela a seguir identifica outras ferramentas que você pode executar usando `command-runner.jar`.

Nome da ferramenta	Descrição
<code>hadoop-streaming</code>	Envia um programa de transmissão do Hadoop. No console e em alguns SDKs, esta é uma etapa de streaming.
<code>hive-script</code>	Executa um script do Hive. No console e em alguns SDKs, esta é uma etapa do Hive.
<code>pig-script</code>	Executa um script do Pig. No console e em alguns SDKs, esta é uma etapa do Pig.
<code>spark-submit</code>	Executa uma aplicação Spark. No console, esta é uma etapa do Spark.

Nome da ferramenta	Descrição
hadoop-lzo	Executa o <a href="#">indexador LZO do Hadoop</a> em um diretório.
s3-dist-cp	Copia de forma distribuída grandes quantidades de dados do Amazon S3 para o HDFS. Para obter mais informações, consulte <a href="#">S3 DistCp (s3-dist-cp)</a> .

# Glossário do AWS

Para obter a terminologia mais recente da AWS, consulte o [glossário da AWS](#) na Referência do Glossário da AWS.



As traduções são geradas por tradução automática. Em caso de conflito entre o conteúdo da tradução e da versão original em inglês, a versão em inglês prevalecerá.